The Emerging Mind

Perspectives on Mental Causation

Annette Wesmann

Master's Thesis in Philosophy (FIL4090)

Department of Philosophy, Classics, History of Art and Ideas

Mentor: Bjørn Ramberg

Spring 2012

2012

The Emerging Mind. Perspectives on Mental Causation

Annette Wesmann

http://www.duo.uio.no/

Trykk: Reprosentralen, Universitetet i Oslo

Abstract:

The subject of this paper is the question of how causally efficacious mental properties can be incorporated into a physicalist framework. The problem is introduced by way of Jaegwon Kim's supervenience-argument, where he formulates it as a question of what causal role is left for the mental to play given certain assumptions of the causal structure of the physical universe. I present his version of physicalism and how mental to physical relations are conceived within this framework. Further, I discuss whether his reductionistic approach to mental causation problem leaves mental properties with adequate causal powers.

In the second half of this paper I consider theories that approach the issue from a non-reductionist perspective, positing emergent causal forces fundamented on the defining characteristics of complex systems. After elaborating on these basic features of such systems I introduce the notion of cognition in its minimal sense as a process of relation between a living system and its environment. This notion is then traced upwards through the levels of evolutionary complexity, reaching the sphere of human agency and cognition. The assumption underlying the theories presented in the second half is that life and mind are fundamentally interconnected and that the causal features of human mentality can only be adequately understood as the pinnacle of a hierarchy of organizational complexity. The downward structuring influences of these cognitive capacities are essential to the process of living and play an indispensable causal role in the characterization of living organisms as self-organizing structures of processes. The claim is that mental notions such as normativity and intentionality cannot be satisfactorily grasped apart from the natural world they emerge out of.

Lastly, I briefly consider the theories of the emergent causal powers of mind presented in the second half in relation to Kim's perspective on physicalism. Here I also discuss whether physicalism has room for higher-order mental capacities as pictured by these non-reductionist theories.

Table of Contents:

SECTION FOUR:

Introduction; stating the issue at hand

The question concerning how mind and body are related to one another has had a central place in philosophical discussion for centuries, and many prominent thinkers have constructed their theories around this issue. From Descartes substance-dualism, postulating mind and body as two radically distinct entities connected solely by the diminutive pineal gland to contemporary reductive identity theories that consider the two to be equivalents, few subjects have garnered more interest and sparked more discussion in the philosophical realm than the question of how our minds and our bodies are connected and interact. The issue can be more broadly stated as how to understand the relation between the mental and the physical, and the answers have implications for a wide range of phenomena central to our existence. It goes to the very core of how we are to understand ourselves as human beings and how, or if, we are able to be genuine causal players in the world on account of our rational agency. This concerns as well the possibility of morality, freedom of the will and of course the underlying subject of identity and selfhood;

> First and foremost, the possibility of human agency, and hence our moral practice, evidently requires that our mental states have causal effects in the physical world. In voluntary actions our beliefs and desires, our intentions and decisions, must somehow cause our limbs to move in appropriate ways, thereby causing the objects around us to be rearranged (…) Secondly, the possibility of human knowledge presupposes the reality of mental causation: perception, our sole window to the world, requires the causation of perceptual experiences and beliefs by objects and events around us.[1]

The above quote presents just a few of the subjects that have the issue of mental causation at its core. The question is most commonly expressed as how mental states can have effects in a physical world and this formulation reveals the paradigm that the current discussion takes place within. Contemporary participants predominantly subscribe to the view that the physical has ontological priority over the mental so that

[1] Kim (2005) page 9

the burden of proof must lay on the causal powers of the mental. The contention is that the natures of the mental and the physical are so diverse that is difficult to see how they can stand in a casual relation to one another, and if they must be chosen between, then the causal powers of the mental must cede to those of the physical.

The belief that what exists in the world is first and foremost of a physical nature is termed physicalism. The different interpretations of this notion will be of central importance in this paper, so for now suffice it to say that the interpretation of physicalism one adheres to sets the parameters for the discussion on what the relation of the mental to the physical consists of. Regardless of which understanding of physicalism one adheres to the concept of existence is most often interpreted in causal terms, owing to the common sense belief expressed in what is know as Alexander's dictum. This expresses the idea that it would be difficult to grasp something that has no effect on anything else in the world, implying that being real and having causal powers go hand in hand.[2] To deprive the mental of causal potency is in effect to deprive it of its reality.[3] In this regard the reality of mental causation is equivalent to the reality of the mental itself and therefore making sense of the causal powers the mental in a physically acceptable manner is of paramount importance if one wishes to argue for its existence. The debate as presented in this paper concerns how much causal reality one can justifiably ascribe to the mental without leaving the physical realm, and the two opponents here are termed reductionistic and non-reductionistic physicalists, or emergentists.

Samuel Alexander was among the most prominent early advocates of the doctrine of emergentism, which initially posited itself as a third way between the extremes of mechanistic reductionisms and dualist theories of mind. The first of the two believed all existents to be ultimately describable in terms of a small number of fundamental laws concerning the paring relations of sets of basic particles[4] whereas the second posited vital entelechies or immaterial forces to describe the workings of the mind. Alexander and his contemporaries were inspired by the ideas of John Steward Mill and G. H. Lewes, who had drawn a distinction between resultant and emergent

[2] Kim (1992) page 133
[3] Kim (2008) page 442
[4] McLaughlin (1992) page 75

properties, whereas the latter distinguishes itself by being irreducible to the sum of its components.[5] In the theories of the early emergentists this idea was expressed through the coming into being of complex systems, where the causal novelty and the related nondeducability of the properties of such systems differentiated them from those whose behaviour could be unproblematically understood by way of the laws governing its parts in isolation. The primary examples of such higher level forces offered of such emergent properties were taken from the realm of chemical and the biological [6] and it was the discovery of lower-level laws explaining the behaviour of what they believed to be emergent phenomena in these domains that to a large degree caused their downfall. It was argued that apparently non-reducible phenomena were the consequence of insufficient knowledge rather than genuine existents. This argument attacked a central weakness of early emergentist theories, namely their inability of the properties they postulated to hold up against increased scientific knowledge. Related to this was their lack of ability to explain the principles structuring such emergent phenomena in a scientifically respectable way, rather Alexander expected the emergent qualities he posited to be accepted "with natural piety".[7]

The positivist school that critiqued early emergentism considered the criterion of any existent claiming scientific respectability to be the availability of publicly, physically testable verification conditions. Out of the positivist school was born the deductive-nomological paradigm of science, fronted most prominently by Hempel and Oppenheim, for whom scientific explanation was a matter of deduction from initial conditions and natural laws.[8] For them any appearance of emergent phenomena was merely indicative of the scope of our knowledge at that time, of having not yet discovered the appropriate natural law.[9] In concert with this view of science Ernest Nagel formulated the pervasive idea that reduction must first and foremost concern the reduction of one theory to another through the formulation of bridge laws that provide connections between the vocabularies of the higher-order theory and that of

[5] Stephan (1992) page 27-28
[6] McLaughlin (1992) page 55
[7] Kim (2005) page 103
[8] Walter (2009) page 98
[9] Stephan (1992) page 38

the lower.[10] Within this paradigm, and inspired by the success of its utilization in physics, reductionistic theories flourished. In the domain of mind-body relations most notably the identity theory of Smart and Feigl, who postulated mental events as type-identical with events in the brain, meaning that a certain mental event "had always and everywhere the neurophysiological characterization initially assigned to it".[11] While the argument from multiple realizability, that identical mental events could be instantiated differently by systems of different material composition (and so could not be tied to certain material structures) brought down such all-encompassing identities, Nagel-type reductions suffered under the accusations that they did not explain what was really at interest, namely the bridge laws themselves (in addition it was argued that it was not genuine reduction because bridge laws demand an expansion of the reduction-base). Today reduction is on the defence while emergentism, now commonly termed non-reductive materialism, is again in favour.[12] [13] Contemporary emergentism has abandoned the historical focus present in the early emergentists' notion of "emergent evolutionism" but retained its core ontological and methodological doctrines concerning the organization of phenomena into autonomous emergent levels and the relationship between basic physics and the special sciences.[14] At its basis is the idea that certain properties at so called higher levels can rightly claim a genuine causal standing in relations to the goings on at lower, unproblematic-ally physical levels. Or, as Jaegwon Kim puts it: "emergent properties are only novel editions to the ontology of the world if they bring with them genuinely new causal powers (this kind of power can only be downwards)".[15] Of those higher-level properties what is of issue here are those that have the characteristics of mentality, and as I hope to make clear in the following section it is their downward causal efficacy related to the physical universe that is the most pressing concern. The question of how the mental can exert causal powers in a physical world is termed the problem of mental causation, which in reality is a collection of issues. It concerns the roles of reasons and beliefs in directing human behaviour, how such roles can be understood in neurobiological terms and why such mental notions matter at all. These

[10] McLaughlin (1992) page 83
[11] Lycan and Prinz (2008) page 6
[12] Kim (1999) page 4
[13] The terms will be used interchangingly in this paper, although I am aware that emergentism is sometimes taken to include elements that are not compatible with non-reductive physicalism.
[14] Kim (1999) page 5
[15] kim (1999) page 25

issues are the subject of a wide array of approaches and for the sake of brevity the treatment of mental causation shall be restricted here to so as to avoid discussions concerning the specific content of mental notions as well as a general analysis of the concept of causation.

In this paper I aim to present an approach to the issue of mental causal powers that offers a positive account of how the mental can exert causal influence in its own right and therefore deserves to be recognized as a genuine existent in the world, rather than an unexplainable by-product of the goings on of the strictly physical level. In order to achieve this I will first and foremost utilize the explanatory resources of scientifically grounded theories rather than leaning on strictly metaphysical treatments of causation and its related notions. The rationale behind this is that the metaphysical approach to the subject to a great degree is fundamented on the methodologies of science and that if we wish to satisfactory grasp these matters then the assumptions underlying our metaphysical doctrines should be brought to light.[16] Appealing to science to understand specifically philosophical questions is termed naturalism, described in the following;

> Another significant mark of philosophical naturalism is a healthy distrust in arm chair philosophy and a priori arguments. To avoid misunderstanding, this does not mean that logical arguments are of little value. It just means that in the end nature is the instance against which we must test the truth or falsity of our theories and the validity of our arguments, not pure rationality, not even purely *logical* thinking."[17]

I will start of by introducing the problem of mental causation as it is concieved of by Jaegwon Kim, after a brief presentation of the central elements of non-reductive physicalism. Thereafter I will consider Kim's proposal for a solution to the problem in light of his understanding of physicalism and state the challenges that such an approach is met by. Using this as my point of departure I then introduce an alternative approach to higher-order causal powers that considers systems and their component

[16] I am aware that it goes both way, that the methods of science are influenced by our metaphysical frameworks as well.
[17] Walter (2009) page 150-151

processes rather than concrete entities and intrinsic relations as the existences of interest. This approach lays the fundament for a theory which posits cognition in its minimal form as an essential element in the self-maintaing processes common to all living beings. The causal power at work here will then be considered in light of its utility in biology, and finally the characteristics describing the activity of living systems will be expanded to capture the defining features of the human mind, in an account of how the human mind emerges from its biological precursors and the fundamental characteristics of life. Lastly I discuss the non-reductive theories presented in this paper in relation to the assumptions underlying Kim's reductive view of physicalism.

SECTION ONE

1.1 Non-reduction presented

The concept of supervenience is the most popular way of expressing the physicalist commitment of the dependence of the mental upon the physical, the degree of dependence between the two being specified by the kind of theory it is utilized within. Donald Davidson was the first to translate the notion of supervenience from its original form as a technical term in ethics[18] by developing the idea that supervenient mental properties are in some substantive sense dependent on or determined by their subvenient base, in combination with the claim that they are not reducible to them in an eliminative sense[19]. Davidson did not further develop the notion of supervenience in his work[20] but Jaegwon Kim, who is one of the most prominent participants in the discussion of the mind-body problem, has been central in exploring the implications of supervenient dependence. The purpose of his exploration has been to make clear

[18] Here it was used to describe the idea that although moral properties cannot be defined in terms of natural properties, they strongly depend on such properties in the sense that if two people are exactly alike in all their natural properties they must be be alike with respect to their moral properties as well. (Beckermann (1992) page 95)

[19] Kim (1989) page 39

[20] It was introduced as an explanatory tool in his theory of anomalous monism, where he postulated that there could not be any laws connecting mental events with physical events and that therefore, given that all events that stand in causal relations must instantiate a law, they are physical events. (Kim (1998) page 33)

the apparently unavoidable inconsistencies of adhering to dependence without reduction, (that is, to be a proponent of non-reductive physicalism) summarized in the following three ideas[21]:

- The central element of supervenience as a physicalist thesis is that it contains a certain degree of dependence on physical properties. It grounds each mental phenomenon within the physical domain by providing a set of conditions that are at least nomologically sufficient for it and on which its occurrence depends.[22] What happens in our mental life is in some way dependent on and determined by, what happens with our bodily processes,[23] so that nothing more than the goings-on at the physical level is needed in order to explain why mental properties come to be instantiated. This dependence-relation is stated by Kim in the following;

  (...) if any system s instantiates a mental property M at t, there necessarily exists a physical property P such that s instantiates P at t, and necessarily anything instantiating P at any time instantiates M at that time.[24]

  This synchrony is expressed as property-covariation. In other words, when two systems instantiate the same P, they can not display differing M's, so that two people in the same neurological state must necessarily exhibit the same intentional content**.** Supervenience as such does not say anything about what this relation consists in, such elaboration is left up to the individual theories, but it does claim that it is necessarily a synchronous relation, meaning that the supervenient event and its base-event begin and end simultaneously,[25] or in other words that they are instantiations of properties of the same objects at the same time. It is also expresses an asymmetric relation because the dependence goes only one way, giving primacy to the physical domain and its laws.[26]
- The non-reductionistic or emergentist element is contained in the idea that although mental properties are dependent on their supervenience-base, they can

[21] Kim (2005) page 33
[22] Kim (1998) page 40
[23] Kim (2005) page 14
[24] Ibid. page 33
[25] Ibid. page 34
[26] Kim (1998) page 7

not be reductively identified with this physical substrate. The supervenience-relation itself is neutral on whether the properties that come about as a result of appropriate lower-level conditions are genuinely emergent or merely "resultant"[27] and therefore unproblematically reducible[28], but non-reductive physicalists claim that certain supervenient property-instantiations are emergent in the sense that they can not be reductively identified with this physical base. A crucial question is of course whether such properties really exist or if they are not all of the latter kind and therefore can be reduced to their subvenient base. Related is the question of what exactly it is that should be reducible to what, and much of the discussion pivots on how this is understood.

- As mentioned at the onset, events are often conceptualized in terms of their causal roles, so that any event or property-instantiation that is to claim genuine existence must be able to show legitimate causal standing. If mental properties are to be non-reducible that means that they must have causal powers that are not derived from their physical realizers in any eliminative sense. They must be definable in a causally autonomous way, all the while avoiding becoming severed from the physical domain and lapsing into dualism. The demands of physicalism are commonly expressed through two principles which are the subject of the next section, and the restrictions imposed by these bring forth the question which forms Kim's argument against the existence of mental events as entailed by the three points above; *"Given that every physical event that has a cause has a physical cause, how is a mental cause also possible?"*[29]

1.2 Closure and exclusion: the principles of physicalism.

Kim's contention is that adhering to what he considers to be the two central principles of physicalism means giving up on the type of causal powers that the non-reductionistic attach to the mental. The first of these principles concerns the assumption that the causal chains leading up to the relevant events can and must be

[27] This distinction was introduced by the early emergentists to differentiate between emergent properties and those that are deducible from the laws governing their parts.
[28] Kim (1999) page 21
[29] Kim (1998) page 38

purely physical. He formulates this in several ways, from its strongest version in the following;

> If you pick any physical event and trace out its causal ancestry or posterity that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and the non-physical.[30]

to its weaker version; "If a physical event has a cause that occurs at *t*, it has a physical cause that occurs at *t*.".[31] The principle that constrains the relevant causes within the physical domain principle is termed the causal closure of the physical, where closure reflects the point that causes of physical events must be contained within a closed physical universe.[32] Kim's understanding of this principle is intimately tied to his belief in the causally and explanatory self-sufficiency of physics; that all physically respectable phenomena can be described, at least as a theoretical possibility, by physics and so there is no need to go outside this domain in order to find a cause of any physical event;[33] **"**On the overall shape and makeup of the world in essential outlines, we must depend on what physics, our fundamental science, tells us".[34] Accordingly, the notion of a causally closed physical domain is very closely related to a certain conception of the structure of the world and the explanatory powers of physics,[35] where the belief is that; "it's only when we reach the fundamental level of microphysics that we are likely to get a causally closed domain."[36] Kim expresses his notion of causality in what he calls robust or productive terms and by this he means the efficient causation involved in energy transfers, momentum and the like that microphysics is concerned with.[37] The reason for his belief that the domain concerned with this kind of causation must be closed is based on the laws of conservation of

[30] Kim (1998) page 40
[31] Kim (2005) page 43
[32] Neither of the two versions offers an unproblematic way of understanding causal closure. In its strongest version the principle begs the question against non-reductive accounts of mental causation because it follows from it that there can not be any non-physical events that are causally related to P-events. The weaker version on the other hand allows for scenarios where a p-event has two non-sufficient causes, one mental and one physical, that together make up a sufficient cause
[33] Kim (2005) page 16
[34] Kim (2005) page 149
[35] The principle of completability of physics, the idea that a complete and comprehensive physical theory of all physical phenomena is at least theoretically possible (Kim (1998) page 40 mind)
[36] Strand (2007) page 58
[37] Kim (2005) page 47

energy and momentum,[38] although these underlying physical laws are not explicitly stated in his works. The law of conservation states that;

> Every physical system is conservative or is part of a larger system that is conservative, where a system is conservative if its total amount of energy and linear momentum can be redistributed, but not altered in amount, by changes that happen in it.[39]

Since Kim considers causation in terms of the transfer of energy or momentum, his assumption is that if 'external' causes where to contribute energy or momentum to a closed physical system, this would violate the conservation-principle. Only fundamental physical forces, intrinsic to the system concerned are allotted causal powers here. His understanding of causal closure has important implications for how he interprets the supervenience-relation between so called higher and lower properties, and I will return to this after an explication of his argument against the credibility of non-reductive physicalism.

The second principle utilized is that of causal exclusion, which states that; "If an event *e* has a sufficient cause *c* at *t,* no event at *t* distinct from *c* can be a cause of *e* (unless this is a genuine case of causal overdetermination)"[40] or, in another formulation: "No single event can have more than one sufficient cause occurring at any given time (-unless it is a genuine case of causal overdetermination)".[41] While the previous principle gives the rationale for choosing P over M as the cause of a physical event, this principle gives grounds for why there must be choosing in the first place. Although cases of overdetermination can be argued to be frequent in the everyday world,[42] they concern causes that are sufficiently independent so as to be outside the domain of the principle. The relation of supervenience between M and P entail that they are not distinct in the manner demanded in order to be considered on the same terms as cases of "ordinary overdetermination" and it is rather the relation between the two causal candidates that demands explanation. The use the causal exclusion-

[38] Gibb (2010) page 363
[39] Ibid. page 370
[40] Kim (2005) page 17
[41] Ibid. page 42
[42] Anders Strand presents an example of a tabletop remaining 92 cm above the floor in virtue of four legs, but where removing one of these would still be sufficient to cause the tabletop to remain 92 cm off the floor. Here causal sufficiency is preserved throughout certain changes of the cause.(Strand (2007) page 79)

principle can be put to depend on how one interprets distinct and sufficient, and the exploration of these two terms has resulted in a welter of differing interpretations. I will not explore the different understandings of distinct and sufficient causes specifically here as I believe that the technicalities will lead my astray from the issue at hand. What is important here is that postulating two distinct causes occurring at the same time , that both claim to be sufficient for an effect, as the non-reductive supervenience-relation is taken to do, apparently leave us with an overabundance of causes.

Before I proceed to Kim's explication of why mental causes are left redundant when these principles are taken into consideration I would like to briefly state the common conception of the events that do the causing. Events are here considered as property-instantiations, that is, they are triplets of objects, properties and times.[43] According to Kim, it is at the level of events as property-instantiations that the causation takes place; "(…) causation is after all is a relation between property or kind-instances, not between properties or kinds as such."[44] His view of the property-instantiations or events doing the causing is central in the theory he presents as an answer to the mental causation and intimately related to his interpretation of the causal closure principle. For the same reason that I do not wish to go deeper into explicating the notions of sufficient and distinct, I will not be treating the notion of causality in general any further at this point. In a sense the underlying theme of this paper is the competing understandings of causality offered by Kim's reductive account and those non-reductive theories I will be presenting later on and so the content of this notion will hopefully become clear in the process of treating these. For now I shall simply state the interpretation of causality as it is most often pictured in approaches to physicalism that rest on the explanatory power of classical physics;

Causality, on the usual analysis, is a succession of events sufficient to permit micro-deterministic state determination of the system involved. That is, the system is causal if a linear sequence can be traced from event to event leaving no "unfilled" gaps in the chain of events. Linear or causal interactions among contiguous events are micro-deterministic and involve a direct transference of energy from one

[43] Strand (2007) page 86
[44] Kim (2005) page 58

event to the next in the chain. "Billiard ball" interaction is paradigmatic of causal interaction in both respects.[45]

1.3 Kim's supervenience-argument:

As mentioned the debate between non-reductive physicalists and those with a reductionistic perspective pivots on their differing views on the relation between the mental and physical properties of causes of physical events. Kim's supervenience-argument is the most influential version of what is termed the exclusion-argument, where the two principles presented in the previous section are utilized in order to challenge the causal commitments of non-reductive physicalism.[46] His goal is to demonstrate that the premise of psycho-physical irreducibility must be rejected in order remain within the parameters of physicalism[47] as set by the two preceding principles. Kim starts off with the assumption the instances of mental properties cause other mental properties, an assumption that is independent of adherence to any physicalist position but that is implied in the commonsense idea that thoughts cause other thoughts in the sense that there is some sort of relation between for example the belief that there is a sandwich in the fridge and the desire to eat it. Supervenience entails that every such mental instance is dependent on, or realized by an underlying physical occurrence, such that the mental event could not have occurred had not the physical event been present and it is here that physicalism enters the picture.[48] It follows from this that in order to cause a subsequent mental event (M*), the mental event considered as a cause must go the way of the physical realizer of M*, that is, P*;

Given that P* is present on this occasion, M* would have be there no matter what happened before; as M's supervenience-base the instantiation of P* at t in and itself necessitates M*'s occurrence at t. This would be true even if M*'s putative cause M, had not occurred- *unless, that is, the occurrence of M had something to do with the occurrence of P* on this occasion*. [49]

[45] Weimer (1976) page 24
[46] Strand (2007) page 48
[47] Kim (2005) page 65
[48] Ibid. page 41
[49] Ibid. page 39-40

So the supervenience-relation entails that it is not possible to have level-bound causal autonomy; any causal relation at levels higher than the bottom level (L) entails a cross-level, L to L-1 causal relation.[50] Here lies the rationale behind the idea that mental causation is a question of downward causation. The question is then how M can exhibit any causal power in its own right in relation to the occurrence of subsequent and related mental events (M*) and in that regard how it can effect the physical events that determine M*. The problem that here presents itself is brought about by the combined adherence to the irreducibility of causally efficacious supervenient mental properties and to the principles of exclusion and causal closure. According to the first of these principles M and P can not both be considered sufficient causes of P*, and according to the second the physical substrate of M must pre-empted M in being the cause of P* if they must be chosen between. What we end up with on this argument is but one horizontal causal chain being on the physical level, and two vertical dependence-relations between M and P and M* and P*. What appears to be causal relations between the higher-level mental properties give way to casual processes on the 'lower' physical level, and the mental causes of physical events are left redundant by their physical causes. The two mental elements involved in this picture play no role in shaping the causal structure, they only supervene on properties that constitute the structure.[51]

Kim summarizes the argument in saying that; "the assumption of casual exclusion and lower-level causal closure disallow downward causation",[52] and so "putative higher-level causal relations give way to causal processes at a lower level."[53] We are left with two vertical dependence-relations and but one horizontal causal chain, which must necessarily reside at the causally closed P-level. In presenting the tension between the vertical determination and horizontal causation present in the supervenience-relation[54]

[50] Kim (2005) page 40
[51] Ibid. page 45

[52] Ibid. page 44
[53] Ibid.
[54] Kim expresses the tension here; "So we have two purportive determinative relationships orthogonal to each other: vertical micro-macro determination and horizontal past-to-future causal determination." (Kim (2005) page 36)

he wishes to point out that "there are inevitable causal entanglements between different levels, raising all sorts of issues concerning causal closure, competition, and exclusion, and forcing some significant philosophical choices."[55] These are issues he believes that the non-reductionistic physicalists must come to terms with if they are to come up with a positive account of multilevel causation. As stated in the beginning of this paper, being a physicalist means explaining the role of mental properties in relation to those that are characterized physically, and as seen in the supervenience-argument such a causal power must necessarily be downward. Calling the type of causation required of the mental 'downward' relates to a certain way of viewing the structure of the world.

SECTION TWO

2.1 Ontological physicalism and the multilayered world

The term downward causation is related to a conceptualization of mental properties as situated at a higher level or being of a higher order than those purely physical,[56] which rests on the idea of a multilayered model of the world where the bottom level is often considered to consist of elementary particles, "or whatever our best physics is going to tell us are the basic bits of matter of which all material things are composed".[57] Each level above that has certain properties that are characteristic for that level, for example metabolism at the level of cellular or higher biological levels and mental properties such as consciousness at level of higher organisms.[58] This ontological scheme is a common way of conceptualizing the world, shared by many of the participants in the discussion on the mind-body problem but they differ as to the how they view the relation between the different levels and ultimately the fundamental bottom layer, as well as the ontological status ascribed to each. The kind of physicalism one adheres to dictates the framework that these relations are to be

[55] Kim (2005) page 64
[56] Those properties that are unproblematically located within the material-energy world.
[57] Kim (1998) page 15
[58] Ibid. page 16

structured by. The central point of emergentism is that although each level is structured out of material particles, the substance of each level being wholly composed of the lower levels and ultimately kinds of elementary particles, the laws governing these lower levels are not sufficient to capture the goings on at higher levels.[59] [60]

Kim's ontological physicalism pictures the hierarchical structure as based on an ordering relation that is mereological, which on his account means that that entities belonging to any level above the very bottom, defined as the level where the entities have no physically significant proper parts,[61] has exhaustive decomposition into entities belonging to the lower levels.[62] He defines his ontological physicalism as; "The view that bits of matter and their aggregates in space-time exhaust the contents of the world."[63] On this ontological scheme the entities belonging to the bottom level have priority over those at any level above, and his belief that such higher-level entities necessarily have exhaustive decomposition entails that they can not be anything but aggregates of parts situated at that level. The idea that any such existent has decomposability as its defining feature is contained in the claim that properties of any existent above the bottom level is micro-structural or micro-based, which means that;

P is a micro-based property just in case P is the property of being completely decomposable into non-overlapping parts, a_1, a_2, …. an, such that $P_1(a_1)$, $P_2(a_2)$….,Pn(an), and $R(a_1 \ldots a_n)$. [64] [65]

Further, he lays out the three defining characteristics of any material object that has proper parts and is therefore decomposable into these. All objects above the bottom level can be described in terms of the basic particles that constitute it, the intrinsic

[59] McLaughlin (1992) page 50
[60] "these laws "emerge" from the laws governing the lower levels of complexity and boundary conditions involving the arrangements of particles, and are in no sense derivative from them" (McLaughlin (1992) page 51). I use the term laws in this paper in the loose sense as describing causal relations and will not treat the issue of how laws are to be understood specifically.
[61] Kim (1998) page 15
[62] Ibid.
[63] Kim (2005) page 71
[64] Ibid. page 84
[65] "Each a is a basic particle or an aggregate of basic particles" (Kim (1998) page 114) Basic particles are those not composed of proper parts and so not further decomposable.

properties of these particles, plus the relations that configure these particles into a structure with "substantial unity".[66] Nothing more is required in order to give a full description of any macro-object, and so once these intrinsic characteristics are fixed the properties characterising their behaviour is also fixed. When connected to his belief that entities above a theoretical bottom level has decomposability as its defining feature, the supervenience-relation is expressed as mereological supervenience. This is the doctrine that properties of the whole are fixed by the properties and relations that characterize their parts,[67] and so "a general claim of supervenience then becomes the Democritean atomistic doctrine that the world is the way it is because the micro world is the way it is."[68] Because of his belief that it is at the level of micro-physics that one is most likely to find the closure conditions required of the physical, the fundamental parts (those which cannot be further divided), must have ontological priority over the wholes they as aggregates constitute.

The importance he places on the idea that microphysics can adequately describe all that is to be considered as physical is reflected in his definition of physicalism seen above,[69] and means that the mind-body relation, as an instance of mereological supervenience[70] can be sufficiently captured by the description of the microphysical properties of the subvenient base. He states that he expects most emergentists to accept mereological supervenience, which involves that one must agree to the assumption that the micro-structural or micro-based properties adequately describe also higher-order entities.[71] [72] As I attempt to show later on, it is exactly this belief that complex physical systems such as the living organism are exhaustively described through its microstructural constitution and the intrinsic micro-causal relations that take place there that are challenged by certain non-reductive physicalist theories.

[66] Kim (1999) page 6-7
[67] "The fact that s has P is fixed once the micro-constituents of s and the properties and relations characterizing these constituents are fixed" Kim (1999) page 85
[68] Kim (1999) page 18
[69] His definition of physicalism as; "all things that exist in the world are bits of matter and structures aggregated out of bits of matter, all behaving in accordance with physics" brings out the central role of physics. (Kim (2005) page 7)
[70] Kim (1998) page 18
[71] Kim (1999) page 7
[72] Put in other words: "the microstructure of a system determines its causal/nomic properties" (Kim (1998) page 23)

2.2 The collapsing of levels.

Even though he expresses supervenience as that of all intrinsic properties of an entity at any level higher than the bottom supervening on the total micro-based properties of that entity at the bottom level,[73] the levels he refers to here are not to be understood in terms of a hierarchy of causally autonomous objects and properties. The demand of decomposability entails that there are no genuine "levels" as such over and above that of the bottom, and any talk of higher-level micro-based properties must be understood in accordance with his belief that the only genuinely causally effective properties are at the level described by microphysics. The microstructural definition of macro-objects and mereological supervenience entails that there is nothing more needed to explain the behaviour of the whole than the properties of the basic parts and their relations. In this way, the bottom level is in effect the universal domain wherein all physically describable macro-objects are contained. Such a bottom level becomes universal through the demand of decomposability and more specifically when characterized in causal terms such as the mental, through the reductive identity of such functional properties with their microstructural realizers.[74] In other words, there is not really any collapsing going on as there is not any genuine causation going on above the causally closed level of microphysics;

> So as far as the supervenience argument goes, the bottom level of fundamental particles (assuming that this is the only level that is causally closed) is always the reference physical domain; there is no step-by-step devolution of causal relations from level to level.[75]

Everything that is dealt with in fundamental physics or reducible, in some broad but clear sense, to fundamental properties can be included in the physical domain[76] Kim states, and lays out three conditions for being allowed within this domain. Those existents that are related to basic physical entities as aggregates, micro-based

[73] Kim (2005) page 59
[74] The first-order realizers of such properties are microstructural properties.(Kim (1998) page 82)
[75] Kim (2005) page 66
[76] Ibid. page 157

properties decomposable into these, or second-order properties defined over physical properties can claim physical status.[77] Common to all these is their reductive identity with the entities dealt with in fundamental physics. On account of this the macro-properties and those that realize them are constituents of the same objects at the same time, so they must be situated at the same level of the micro-macro hierarchy.[78] [79] It makes more sense to speak of orders than layers he says,[80] because there is in reality only one genuine level here, that of the bottom, which is "thought to consist of elementary particles, or whatever our best physics are going to tell us are the basic bits of matter out of which all material things are composed".[81] [82]

Kim's understanding of the scope and nature of the microphysical domain sets the stage for a reductionistic explication of mental causation, where he seeks to reductively identify the causal power of the mental with that of its microstructurally defined realizers. At its basis is the ontological priority of the parts of the whole, the parts being fundamentally those which can be explained solely in terms of a causally closed micro-universe. While causation taking place here can be termed macro in the sense that it concerns the object as a whole, the object as a whole is nothing but an aggregate of composite basic particles and their intrinsic relations.

2.3 Non-reductive physicalism again

Non-reductive physicalism on the other hand credit properties of existents on higher levels of complexity with a causal standing that can not be derived from the powers of their constituents, not even taken as an aggregate. In other words they are against the

[77] Kim (1998) page 114-115
[78] Ibid. page 83
[79] "Properties in the reduction-base for an identity based reductionism must be at the same level as the properties to be reduced. This means that the base property must be instantiated by the same objects as the reducible property. Reasonably; if two properties are identical they are instantiated by the same objects" (Strand (2007) page 53)
[80] Kim (1998) page 83
[81] Ibid. page 15
[82] He does however use the term "levels" and "orders" interchangingly in his works, so I will permit myself as well to use the terms loosely in this paper.

idea that mereological supervenience based on the microstructural definition of macro-objects capture the nature of the physical world in its entirety;

When we speak of non-reductive materialism we are considering a position that subscribes to the basic physicalist idea that everything that exists is constituted by matter, there being no "non-physical" forces at play, but that when such material systems reach a certain degree of complexity they begin to exhibit novel properties not reducible to the properties of their constituents.[83]

They stand in opposition to the idea that physics adequately capture the nature of the physical world and consider emergent phenomena at higher levels to be neither predictable nor explainable on the basis of their emergence base, and accordingly they provide additional causal structure to the world not describable solely in terms of physical science. The forces are commonly considered to be 'configurational forces' which come about as organizational complexity increases, not through the paring relations of particles that physics (as pictured by Kim) is concerned with.[84] As stated in my very brief treatment of the concept of causation underlying Kim's notion of physicalism, in physical science causation is often pictured as a linear relation among spatio-temporal particulars in direct "contact" with each other[85]. Essentially the effect is here the sum of its causes, being decomposable into elements which have a one-to-one correspondence.[86] Configurational forces on the other hand are not grasped within this conception of causation, having the characteristics of non-linearity[87] and extrinsicness. It is in terms of these forces that mental causation is formulated in the emergentist accounts related in this paper. According to the non-reductionist mental phenomena have sufficient causal power in a physically relevant sense so as to be considered a substantial existent even within a physical world, which entails that the notion of the physical world must be expanded beyond that which is captured by Kim's ontological physicalism. It is a challenge to the idea expressed by Kim that the world consists nothing but bits of matter described by the laws of physics.

[83] Kim (1999) page 4
[84] McLaughlin (1992) page 53
[85] "It is of the essence of material substances that they have determinate positions in spacetime and that there be a determinate spatiotemporal relation between each pair of them" Kim (2005) page 86
[86] Thompson (2007) page 68
[87] "Such systems allow very small of intial values to lead to very large divergences of future states. Trajectories which are in close neigbhourhood can diverge exponentially" (Stephan (1992) page 33). In other words, the effects cannot be deduced from the sum of its causes in a straightforward way.

As shown in the supervenience-argument the discussion revolves around whether the causal powers credited to higher-level existents on the non-reductionist account actually can claim a substantial place within a physicalist ontological scheme or whether non-reductive physicalism is an oxymoron. Of special concern in this paper is whether the activity of living systems can be adequately described without being reductively identified with the activity of its intrinsic material parts. Kim himself believes that there are only two genuine options if one does not wish to leave physicalism behind; reducibility or epiphenomenalism:

> That is, if you have already made your commitment to a version of physicalism worthy of the name, you must accept the reducibility of the psychological to the physical, or, failing that, you must consider the psychological as falling outside your physically respectable ontology.[88]

The counterclaim of modern non-reductionist physicalists is that; "the laws of physics and chemistry can not account for the specific kinds of living things- that is, the particular configurations of matter in organisms".[89]

2.4 The physical and the mental

Ultimately the goal is to reach an understanding of how the mental is situated within biological life as a physical entity, as the naturalistic commitments require of those theories that I will be utilizing in this pursuit. In this regard it is of course pertinent to understand what is signified by the use of the terms mental and physical. While it its clear that what Kim considers as physical is intimately connected to his notion of decompositionality and therein the relation to the entities treated by the fundamental physical sciences, it is apparently a fuzzier concept when used by non-reductive physicalists. As mentioned above, it is the configurational forces of matter that is of central importance to them. It is important to note that non-reductive physicalists adhere to the general tenet of physicalism that the world is wholly composed by

[88] Kim (1989) page 32
[89] Ayala (1974) page xiii

matter[90] and they do not consider emergent ontological entities but rather emergent levels of causal efficacy.[91] All existents are still physical, but some have characteristics not held by its microstructural relations.

Concerning the mental, a central feature is that of intentionality.[92] Franz Brentano considered the characteristic of intentionality to be the mark of the mental[93] pointing to the capacity of the mind to intend the nonexistent[94], the directedness of mental states being what makes it so difficult to place within a physicalist framework.[95] For how could some physical thing in the head come to bear the aboutness-relation to anything out in the external world?[96] A common answer is that the semantical content of such relations can be identified with some state of the brain that represents the state of affairs instantiated in the world.[97] This suggestion begets a number of problems, such as the mentioned issue of how they can represent things that have no real-world correlate, as well as how to exactly representations are structured out of basic brain-features. Here the concern is that of mental causation, that is, how intentional states such as reasons, desires or intentions cause an action (more generally expressed as how reasons can be causes). They stand apart from the physical processes underlying them on account of their semantic dimension[98] and the relational or extrinsic character of their nature. Related is the notion of normativity, the prescription of value to the semantic content of intentional states or relations, which is as well difficult to capture in a physical world of bits of matter described by laws of nature. Kim puts the problem of intentional causation in terms of supervenience-relation assumed for mental properties. The issue he states is that while the internal cause of physical behaviour must be supervenient on the total internal state of the organism at that time (implying that two people in an identical such state must emit identical motor-output),

[90] Kim (200) page 150
[91] Murphy and Brown (2009) page 8
[92] Kim includes states such as belief, desire, memory and perception in the class of intentional or cognititve properties (Kim (2005) page 162), as well as perception, information processing, inference and reasoning, and using information to guide behaviour (Kim (1999) page 9)
[93] Walter (2009) page 29
[94] Murphy and Brown (2009) page 155
[95] We saw earlier that Kim believes all causation to concern intrinsic properties or relations, so it becomes difficult to incorporate the apparent extrinsic or relational character of intentional causation here.
[96] Lycan and Prinz (2008) page 87
[97] Ibid. page 87
[98] What it is directed at is its semantic content

having identical intrinsic properties does not necessarily mean that two people will instantiate the same semantical properties. This is because the instantiation of a semantic property is essentially a relational fact involving details about the organism's history and its environmental conditions.[99] If, as the supervenience-argument seems to make unavoidable, all causation must involve physically respectable, intrinsic properties of the system; "How can extrinsic, relational properties be causally efficacious in behaviour production?".[100]

2.5 Aspects of reductionism

Since the central point of contention between the two approaches to physicalism outlined above is reduction it will be useful to present an overview of the different elements contained in this notion. Nancy Murphy and Warren Browns account offers a clear and precise summary which I will utilize for this purpose. They divide the term into five related elements, which are as follows; (partly quoted)[101]

1. **Methodological reductionism**: a research strategy for analyzing the thing to be studied in terms of its parts. Early modern science considered this the only legitimate approach to any scientific subject of study and it is the preferred research strategy for those that adhere to the belief that the understanding of all entities can only come from the understanding of its physio-chemical parts[102]. This research-strategy dictates certain limitations on scientific research, such that the only acceptable evidence is that which is definable and defensible in terms of physical science and that the pursuit of this evidence should be conducted on simple and relatively isolated systems, assuming that the behaviour of large and complex systems is nothing but or equals the sum total of the behaviour of its constituent parts.[103] As I will attempt to show later on in this paper, the recognition of the difficulty in understanding specifically biological concepts, and hereunder the as the notions of cognition and

[99] Kim (1998) page 36
[100] Ibid. page 37
[101] Murphy and Brown (2009) page 47-48
[102] Ayala (1974) page viii
[103] Skolimowski (1974) page 207-208

intentionality, has demanded a complementary approach to this tradition methodology of science.[104]

2. **Epistemological reductionism**: the view that laws or theories pertaining to the higher-levels of the hierarchy of the sciences can (and should) be shown to follow from lower level laws, and ultimately from the laws of physics. So this aspect of reductionism is a related to the previous, which dictates research-strategies. In the period dominated by the deductive-nomological model of science the search for such underlying laws was the rule. Within this paradigm Ernest Nagel developed what became the most pervasive way of considering reduction. Nagel was more concerned with the reduction of theories than with the phenomena within these, and this reduction was formulated as the logical or mathematical derivation of the laws of the theory targeted for reduction from the laws of the base theory.[105] This required bridge-laws connecting the predicates of the two theories as their vocabularies were disjoint, but these bridge-laws themselves were left unexplained. Given the widely accepted nomological requirement on causal relations, namely the condition that "events standing in a causal relation must instantiate a causal law"[106] then in a broader sense the formulation of reduction in terms of laws is still relevant;

> The general question is whether the theories an experimental laws formulated in one field of science can be shown to be special cases of theories and laws formulated in some other branch of science. If such is the case, the former branch of science is said to have been reduced to the latter. This is the sense in which philosophers of science most often discuss questions of reduction.[107]

3. **Causal reductionism**: the view that the behaviour of the parts of a system (ultimately, the parts studied by subatomic physics) is determinative of the behaviour of all higher-level entities. Thus, this is the thesis that all causation in the hierarchy is 'bottom-up'. The behaviour of an entity is determined by

[104] As I will bring up later, the defining characteristic of living systems is that they are open, whereas science such as physics has traditionally concerned itself with closed systems where the causes can be viewed unproblematically as linear and intrinsic.
[105] Kim (2005) page 98
[106] Kim (1998) page 33
[107] Ayala (1974) page ix

the behaviour (or the laws governing the behaviour) of its parts.[108] This kind of reductionism comes to expression in Kim's notion of mereological supervenience and it is this aspect that is the central element of discussion in this paper.

4. **Ontological reductionism**: the view that higher-level entities are nothing but the sum of their parts. This position thesis is further divided into two distinct positions:

   4a) When one goes up the hierarchy of levels, no new kinds of non-physical "ingredients" need to be added to produce higher-level entities from lower. Differences on this position are what distinguish physicalism from dualist thesis that posits "vital forces" or "entelechies" to explain how living beings arise from non-living materials. In order to be within the physicalist realm one must claim adherence to this version of ontological reductionism, but this does not necessarily imply that one must submit to the following aspect of reductionism.

   4b) "atomist reductionism": The claim that only the entities at the lowest level are *really* real; higher-level entities- molecules, cells, and organisms- are only composite structures (temporary aggregates) made out of atoms. The entities at the bottom of the ontological scale have priority over the things they constitute.[109] Kim's view of macro-properties as being mereological aggregates of their base-properties, exhaustively described in terms of its basic particles, their intrinsic properties as well as the relations between them, implies that he belongs here.

As well as the term physicalism and causation, there is no general agreement as to how reduction is to be understood but this general way of dividing the different elements contained in the term capture the main points of reductionism. Given that many believe existents must be considered in terms of their causal standing, ontological reductionism is intimately tied to that of causal reductionism, which in turn relates to epistemological reductionism through the requirement that causal

[108] Murphy and Brown (2009) page 62
[109] Murphy and Brown (2009) page 47-48

relations are formulated in terms of laws or expressed as such. Lastly epistemological reductionism concerns methodological reductionism as it is this that dictates the research-strategies for formulating these laws. In this paper I will not be concerned with strictly differentiating the diverse aspects, as considering causal reductionism inevitably will have epistemological and methodological connotations.

In addition to the four elements laid out above, there are two main ways of utilizing reduction. The first is exemplified in the type identity theory of Smart and Feigl's mind-brain identity thesis,[110] which was briefly mentioned in the introduction. On the other hand we have token reductionism which does not aim for the all encompassing reductionism entailed by the type-version. By limiting the reduction of a given higher-order element to certain contexts[111] this kind of reduction is able to resists the multiple realization arguments levelled against type-identity reductions to a greater degree. It is also a kind of reductionism that all physicalists can safely admit to, given that physicalism in its broadest sense states that all concrete entities are physical entities. On Kim's token physical reductionism it is the local reduction of mental events to their physical realizers that is of interest. Instantiations of mental properties at a certain time are reducible to instantiations of their micro-particulars at that time, and this reduction is relative to the species or structure under consideration.[112] Concerning the possibility of reduction of mental properties to their physical realizers, Kim believes the best solution is through functionalization.

2.6 Kim's functionalism:

Kim's suggestion for a solution to the problem of mental causation posed by the supervenience-argument is to consider mental properties as second-order functional properties with physical realizers,[113] and by this providing a positive account of how the two are related. For Kim, reduction is essentially functionalization and it is the functionalization of mental properties that enable them to escape the supervenience-

[110] Kim (2005) page 14
[111] Understood as "(…)the local reducability of psychology, local relative to species or physical structure-types" (Kim (1989) page 39)
[112] Kim (1998) page 110
[113] Kim (1998) page 24

argument.[114] By positing that instantiations of any mental property is reductively identifiable with the synchronous instantiation of some physical property that satisfies the causal specifications of the mental property he believes that his commitment to causal closure can be reconciled with what he considers to be robust mental causation.[115] Through functional reduction the mental can be preserved as part of the physical domain while not being eliminated as a genuine existent. He views his functional reductionism as 'conservative' rather than 'eliminative', meaning that M is not something "over and above" P in the emergentist sense while still being conserved to a certain degree. He believes that this type of reduction requires identities, because; "to conserve X as Y means that X is Y".[116] According to this approach mental properties are second order properties that consist in having other properties that have certain relations to one another, stated as;

F is a *second-order property* over a set **B** of base (or first-order) properties iff F is the property of having some property P in **B** such that D(P), where D specifies a condition on members of **B**.[117]

The functional reduction of a higher-order property E to the domain B of properties serving as the reduction base is achieved by way of a three step process. First E must be defined by its causal relations to properties in the reduction base B, which can be stated as such:

Having E= def Having some property P in B such that (i) C1 ,…. Cn cause P to be instantiated, and (ii) P causes F1 ,…, Fm to be instantiated. (We allow either (i) or (ii) to be empty).[118].[119]

The aim is to make E relational (as in non-intrinsic) with respect to other properties in B. The E's being instantiated is for a property P to be instantiated, this instantiation bearing causal or nomic relations to the instantiations of a specific set of properties in the base domain.[120]. Kim uses the functional reduction of the gene to DNA molecules as an example. Here the gene is first defined in terms of the casual work that the

[114] Ibid. page 116
[115] Strand (2007) page 50
[116] Kim (1999) page 15
[117] Kim (1998) page 20

[118] Kim (1999) page 10
[119] A simpler way of expressing this is: "having M = (def) having some property or other P (in the reduction base domain) such that P performs causal task C" (Kim (2005) page 101)
[120] Kim (1999) page 11

property is to perform, that is, being a mechanism that encodes and transmits genetic information[121]. This is a purely conceptual step, simply reinterpreting the property to be reduced in terms of the causal work it is to perform. It does not involve any empirical or factual information about the property subject to functionalization,[122] just establishing a conceptual or definitional connection between E and the selected causal role.[123] According to Kim, this step in and of itself is necessary and sufficient for functional reducibility. Whether the property in question can be reduced or not, that is, if there are actual lower-level mechanisms there that perform the assigned causal work is an independent issue that is a matter of scientific research.[124]

The second step is to identify the properties serving as the realizers of E in B, the properties that in the reduction base that perform the specified causal task. Whether a given property realizes a functional property is an empirical, contingent matter, and in order to avoid multiple realization worries they are restricted to certain contexts. As long as the realizing base-property stands in the right causal or nomological relations to other properties it can serve as its realizer or, as Kim puts it; "Any mechanism that gets activated by the right input and that, when activated, triggers the right response serves as a realizer of a psychological capacity or function".[125]

The second order properties are defined in terms of the causal task that is to be performed by the first-order realizer of that property and since these functions can be performed by different physical structures in different systems the realizers can be multiply realized, that is, the reductions are local. In systems with similar microstructures however the realization relation must be equal because, as mereological supervenience requires; "once the system's physical constitution and the prevailing laws of nature are fixed, that fixes whether or not P realizes M in that system".[126] So, different realizers of the same functional property can be found in

[121] Kim (2005) page 101

[122] Ibid. page 28
[123] Kim (1999) page 11
[124] Kim (2005) page 164-165
[125] Kim (1998) page 21
[126] Ibid. page 23

different organisms, and even in the same person over the span of a lifetime.[127] Science has discovered that in terrestrial organisms the DNA molecules are the mechanism that performs the task of encoding and transmitting genetic information. Whether this realizer of the gene can be generalized to species of other worlds is not of interest here, unless such species where to become the subject of research. Unlike reductions that are postulated as concerning all possible worlds, this type of reduction will only be as broad as science specifies it. In other words, it is nomologically necessary but metaphysically contingent.

Thirdly, a theory must be produced to explain how realizers of E perform the causal task that is constitutive of E. These two steps are not independent of each other. As with step two, this is a matter for the relevant sciences, in the case of the gene and DNA molecules it is up to molecular biology to provide an answer to how DNA is able to perform the causal task specified by the gene.

On the functional model the causal power of every instance of some mental property is nothing over and above that of its realizer on that occasion.[128] In essence it identifies the functional M property with the P-property that realizes it.[129] [130] However, Kim claims that the M to M* causation remains very genuine and real, in virtue of being the very same causation as P->P*. The reduction collapses the two levels into one and there is here one causal relation, not two.[131] This reflects his answer to the supervenience-argument in terms of microstructural decomposability, where the properties of macro-objects by their constitution of decomposability is identical to an aggregate of their micro-particulars, their causal power being the sum of that of these.[132] While the all-encompassing Nagelian bridge laws which have been

[127] This accords well with the discovery of the plasticity of the brain, where research has shown that the very same populations of neurons can take on different roles depending on changes in the overall makeup of the brain.
[128] The notion that every functionalizable property has exactly the power of its realizer on that occation is termed the causal inheritance-principle (Kim (1999) page 16)
[129] Strand (2007) page 51
[130] Note that the actual identities are a posteriory identities, not identities by conceptual analysis or definition. (Kim (2005) page 113)
[131] Kim (2005) page 54
[132] I interpert the statement that; ”Whether a given property P is a realizer of a mental property M depends on the nature of the system in which P is embedded, since in psychology the input-output behaviour of the total system is what is of concern, and the causal role that P plays will depend on the makeup (“causal wiring”) of the system as a whole” Kim (1998) page 22) to be the functional equivalent to the microstructural characterization of the properties of macro-objects in general. He does

the reference point for much of the discussion on reduction suffered under the arguments of multiple realizations, Kim believes that moving from bridge-law reduction to identity reduction by way of functionalization has made reduction less vulnerable to this kind of argument. Because it is the second-order property can have different realizers dependent on the system it is related to, the concept itself will be as heterogeneous as the heterogeneity of its diverse realizers.[133] The search for the specific realizers is only concerned with a limited number of systems, and does not attempt generalizations beyond that which is scientifically responsible.

It is important to note that, in line with his view that all causation takes place at the level of instances, Kim differentiates the higher-order property itself and the instances of that property. It is the instances of such properties that are treated as functional reducible;

> (…) system s has E, in virtue of s's instantiating one of its realizers, say Q. Now s's having E on this occasion just is its having some property meeting causal specification C, and in this particular instance s has Q, where Q meets specification C. Thus s's having E on this occasion is identical with its having Q on that occasion. There is no fact of the matter about s's having E on this occasion over and above s's having Q. Each instance of E, therefore, is an instance of one of E's realizers, and all instances of E can be partitioned into Q_1-instances, Q_2-instances,…, where the Q's are E's realizers. Hence the E-instances reduce to Q_i-instances.[134]

While he believes that the instances of higher-order properties retain their causal power in virtue of being conservatively reduced to their realizers, he suggest eliminating E as a genuine property and only recognizing the concept or expression E, devoid of causal powers of its own. Given the multiple realizations of its instances E becomes causal or nomologically heterogeneous, making it unfit to figure in laws and therefore disqualified as a useful scientific property.[135] The second-order concepts designate the first-order properties in he reduction base disjunctively.[136] Anders

---

state that the first-order realizers of second-order functional properties are micro-structural properties in Kim (1998) page 82.

[133] Kim (2005) page 26

[134] Kim (1998) page 15

[135] Ibid. page 17-18

[136] "On the present view, the concepts introduced by second-order designators pick out first-order properties disjunctively". (Kim (1998) page 105)

Strand raises the concern that it is problematic to establish identities between second-order properties and their realizers when this identity-base is viewed as the disjuncts of their first-order realizers.[137] It seems to include causally irrelevant details because the instantiation of a disjunctive property by an object supervenes on a singular disjunct which is causally sufficient for the effect, and accordingly the disjunctive property is causally excluded by its disjuncts.[138] It is a legitimate concern, but I will not go further into an exposition on the understandings of disjunctive here as I consider the problems facing mental properties construed as second-order properties in the next section. Kim does seem to admit that there cannot be any general property-identities in functional reduction[139] in the following quote;

> Now, any instance of M must be ether a P1 instance or a P2 instance, and this means that instances of M are not going to show the kind of causal/nomological homogeneity we expect from a scientific kind. In short, multiply realizable properties are causally and nomologically heterogeneous kinds, and this at bottom is the reason for their inductive unprojectability and ineligibility as causes.[140]

Whereas the instantiation of a particular mental property is functionally reducible to a certain instance of its realizer (relative to the specific structure under consideration, given that it is a local reduction), the specific realizer doing all the causal or explanatory work of M on that occasion, the property as such has no causal role to play of its own.[141] Its role is merely communicative or descriptive;

> When we use the functional characterization of pain (..) we let others know that we are referring to a state with certain input-output properties; a neural characterization of its realizer would in most ordinary contexts be useless and irrelevant.[142]

While the second-order properties themselves are left causally inert and therefore not deserving of the name properties at all, (Kim prefers the term second-order concepts

---

[137] Strand (2007) page 66
[138] Strand (2007) page 68
[139] Kim (1998) page 112
[140] Ibid. page 110

[141] While the property designates a disjunct of its realizers, at a certain occation there is only one realizer doing causal work, or put in explanatory terms; ”what we have is a disjunction of two explainations, not a single disjunctive explaination” (Kim (1998) page 108)
[142] Kim (1998) page 104

or property-designators[143]), their instances provide us with an explanation of the supervenience-thesis; the mental supervene on the physical because instances of mental properties are second-order functional properties with physical realizers (and no non-physical realizers).[144]

2.7 What's left of the mental in Kim's account?

In short Kim's functional reductionism states that mental properties can be explained solely through the discovery of the causal or nomic relations obtaining in the base domain of its realizers.[145] Because Kim adheres to Alexander's dictum stating that for something to exists means that it must be able to make a difference to the world he must prescribe the mental with sufficient causal powers if they are not to be eliminated by their realizers. But according to the causal inheritance principle, when a functional property is instantiated on a given occasion in virtue of one of its realizers being instantiated, then the causal powers of this instance of the functional property is identical with the causal powers of its realizer,[146] and by this the causal powers of mental properties are reduced to the powers of its physical realizers.[147] As mentioned above, the property itself is regarded by Kim as eliminatively reduced, ending up as a designator of its realizers with no causal efficacy apart from the particular instantiations of its realizers. Since all the causal or explanatory work done by an instance of M that occurs in virtue of the instantiation of a specific realizer is done by that realizer, we must ask what, if anything at all justifies us speaking of the mental at all, except as mere designators.[148] It seems that the accusations he levels against Donald Davidson's anomalous monism, namely that the fact that M is a mental event is causally irrelevant[149], can be directed towards his reductive account of higher-order properties as well;

---

[143] Ibid. page 106
[144] Ibid. page 24

[145] Kim (1999) page 14
[146] Ibid. page 16
[147] "(…) an instance of a second-order property cannot have causal powers beyond those of the realizing properties involved" (Kim (1998) page 129n)
[148] Kim says; "only causally relevant or efficiacious properties should count as individuating properties (..) it is highly implausible to say that events that are indiscernible in respect of causal properties can yet be distinct events (–so the M instance just is the P instance)" (Kim (1998) page 56)
[149] Kim (1989) page 34-35

Event *m's* causal relations are fixed, wholly and exclusively, by the totality of its physical properties, and there is in this picture no causal work that *m*'s mental properties can, or need to, contribute.[150]

Summarizing the consequences of his main positions, the demand of decomposability of higher-order properties means that the microstructure of a system determines its causal or nomic properties,[151] whether they be considered in causal terms as functionable or more generally as macro-objects. As his adherence to mereological supervenience implies, it is the goings-on at a potential bottom level that determines the causal standing of all existents above this. The question is whether such a reduction based identity situated at the lowest conceivable level can satisfactory accommodate mentality within the causal structure of the physical world, as Kim believes that it does.[152] To quote Strand:

What we want from an account of mental causation is an understanding of how mental events are causes *in virtue of their mental properties*. This is not saved by claiming that mental events are causes in virtue of being identical to instances of physical properties.[153]

In other words the reductive identification of such properties drains them of sufficient causal power. In his causal drainage-argument Ned Block expresses the broader worry that positing mental properties as second-order functional properties[154] reductively identifiable with their microstructural realizers will leave us without any substantial causal powers above the fundamental level.[155] The charge is that by identifying instances of micro-based macro-properties with instances of their microstructural properties[156] Kim makes any causation above an eventual bottom level impossible, because, as Strand points out; "he is committed to a supervenience-base at the level of fundamental physics".[157] On Kim's view of the dependence-relation between higher-level micro-based properties and their "lower-level counterparts",[158] all intrinsic

[150] Kim (1998) page 34
[151] Ibid. page 23
[152] Ibid. page 38
[153] Strand (2007) page 66
[154] Kim (1998) page 24
[155] This worry is part of a larger argument, where the question is whether or not the problem generalizes so as to make all special-science properties epiphenomenal (Kim (1998) page 112)
[156] Kim (2005) page 57
[157] Ibid.
[158] Ibid. page 59

properties of O (being a macro-object), at any level higher than L, (levels understood in terms of "modes of decomposition of material objects into physically significant constituents"[159]) supervene on the total micro-based property of O at level L.[160] Stated in another way; "for any object O. O's micro-based properties at level L supervene on O's total micro-based property at level L*, where L*<L".[161]

Concerning mental properties construed functionally, the realizers of such second-order properties must be superveniently dependent on lower-order realizers and so forth until we reach such a stipulated bottom level. They are included in this notion of supervenience in virtue of the first-order realizers they supervene on[162] being micro-structural properties and therefore necessarily being decomposable onto the lowest level.[163] Given the causal inheritance-principle all causal powers must ultimately reside on this level, either directly or through supervenience on intermediate orders or levels. Consequently, if there is no such level and matter is indefinably divisible then there will be no causal powers anywhere, because every member of an infinite series of supervenient properties will have a lower member to which it cedes its causal power, a counterintuitive notion that puts the existence of everything in peril. The threat posed by first-order realizers to pre-empt the causal claims of the second-order properties they realize is a consequence of the supervenience-argument and concerns all second-order properties, whether functional or not.[164] In fact, Kim does not believe that there is a sharp distinction between micro-based and functional properties, it is rather a matter of how they are defined, for as he says "there can surely be micro-based properties some of whose constituent properties are functional properties."[165]

[159] Ibid.
[160] Ibid.
[161] Ibid. page 60

[162] Second-order properties, hence mental properties construed as functional properties , supervene on their first-order realizers
[163] Kim (1998) page 82
[164] Ibid. page 55
[165] Ibid. page 115

2.8 Kim's answer to the drainage-issue:

The answer he gives to Blocks drainage-argument rests on his view that the bottom level must be taken to mean the causally closed fundamental level of microphysics, which through the reductive identity of all higher-level existents with the micro-particulars that constitute them, is in effect the universal domain given that each level includes all mereological aggregates of entities at that level.[166] To reiterate, Kim resists talk of ontological levels such as they are conceptualized by non-reductive physicalists because the closure-condition implies that there can not be talk of any genuinely ontological levels above that described by physics. By making any object or property above this level reductively identifiable, either through microstructural composition or second-order properties defined over physical properties, the closed physical domain will include such existents as well. Accordingly, causal powers can be held by aggregates of basic physical entities, micro-based properties decomposable into these, as well as second-order properties defined over physical properties.[167] Common to all these is that they are reductively identifiable with specific mereological configurations of this basic physical substrate, as the supervenience-argument demands.

If there is no fundamental physical level to serve as the universal domain by way of its causal closure, he points out, then the supervenience argument cannot get a toehold, because the required closure-premise is missing; "a causal collapse to the level below would occur only if the lower level is causally closed".[168] On the other hand, if such a potential 'bottom level' is available to serve as a reduction base for all identity-based reductions, then this reduction will be "the stopper that will plug the cosmic hole through which causal powers might drain away."[169] The idea behind this is that it is the irreducibility of supervenient properties to their subvenient base that makes their causal powers become pre-empted by those of the base-level,[170] and by including both microstructurally and functionally defined macro-entities in the

[166] Kim (2005) page 65
[167] Kim (1998) page 114-115
[168] Kim (2005) page 67
[169] Ibid. page 68
[170] Ibid. page 60

physically closed domain of microphysics the pre-emption problem is resolved. He gives an example of how the reductive identity of water is continued from the commonly accepted 'water=H2O' downwards through that of H2O being a total micro-based property of water at the molecular level. This microstructural composition is further decomposable into the underlying level and so forth[171]. The lowest level being the union of all the micro-levels involved in the chain of identity,[172] this retains the properties of macro-objects through mereological aggregation.

Certainly the reductive identification of any macro-entity with such a union of causal powers relieves the tension between the vertical determination and the horizontal causation apparent in the supervenience-relation,[173] but as I attempted to express previously, the main concern here is not the existence of any causal powers at all but whether the dependence of all 'higher-level' causation on such a bottom level gives sufficient room to claim the existence of mental properties at all in terms of causal powers. As Strand points out, resorting to a reduction base of properties based at the fundamental physical level in order to avoid the drainage-charge means no causation above this eventual fundamental level[174] and as well it makes the macro-entities that supervene on very fragile, the slightest change in its microphysical realizers affecting the constitution of macro-properties**.** In short, Kim's reductive identification of mental properties to those of a closed physical domain can be accused of leaving us without the kind of power required for postulating genuine causal agency. In Kim's own words;

> When we try to relieve the said tension by saying that this instance of M caused P* to be instantiated (thereby causing M* to be instantiated), we must reflect on what this means in terms of the causal inheritance-principle. The principle says that this instance inherits all of its causal powers from the first-order property that realizes M on this occasion, and this first-order property is P. So this instance of M can claim to cause the P*-instance insofar as, and only insofar as, this instance of P causes it. And it becomes entirely unclear what could motivate us to countenance the M-instance, in addition to the P-instance that realizes it, as a cause of the P* instance.[175]

[171] Kim (2005) page 68-69
[172] Ibid. page 69
[173] Ibid. page 38
[174] Strand (2007) page 48
[175] Kim (1998) page 55

2.9 Macro-causality and downward causation

In fact, Kim does not seem too concerned with what is left of higher-order causal powers when such properties are interpreted as functional instances or with the causal power of micro-based macro-properties in general. On the latter he says that although the set of causal powers held by a micro-based property can be explained microstructurally, it does not mean that they are identical to the causal powers of its micro-constituents.[176] These because they are identical to specific mereological configurations involving these micro-properties, not supervene on them taken individually or as a group.[177] He gives an example of a baseball, being a composite structure of micro particles exercising 'macro-causality' by breaking a window. The causal powers of the baseball are not had by any of its constituent micro particles or proper parts although these powers are determined by and explainable in terms of its microstructural features. This does not banish macro-causation out of existence, he says, even though macro-causal relations are constituted by or composed of a bunch of micro-casual relations.[178] So the inclusion of aggregates of micro-particulars in the closed physical domain means that macro-objects are allowed causal powers not had by any one of its constituents.

> Perhaps, macro-causal relations are constituted by, or composed of, a bunch of micro-causal relations: But that does not banish macro-causation out of existence any more than the fact that the baseball is composed of micro particles entails its non-existence.[179]

Concerning second-order functional properties, he states that they are causally heterogeneous but not causally impotent,[180] which I assume points to the fact that every instance of a functionalizable mental event gets its causal power from that of its realizer, because he can not be referring to any causal power of the second-order

---

[176] Kim (1998) page 117
[177] Ibid.
[178] Kim (2005) page 56
[179] Ibid.
[180] Kim (1998) page 116

properties themselves. Given that he states explicitly that these are not even to be considered as genuine properties, have only communicative relevance, mental properties as such can not have any causal powers above their individual realizers; "(…) the causal powers of mental properties turn out to be just those of their physical realizers, and there are no new causal powers brought into the world by mental properties".[181] So it is through the reductive identity to instances of their first-order physical realizers that mental properties construed functionally retain any causal power. "Each instance of M as a realizable property has exactly the causal power of its realizer on that occasion"[182] he says, and the causal heterogeneity of M is given by the multiple realizability of its realizer, depending on the nature of the system that it is realized in. What then, is left of the causal power of the mental? Just that of the power of the realizer it is functionally identified with. Given that their first-order realizers are microstructural properties we are faced with the same problems concerning functionally reducible mental properties as with micro-based macro-properties in general. Kim dismisses this concern rather causally; saying that giving up any distinctive causal role of the mental is a steep price to pay but it is one that we will have to pay if we are to remain within the physicalist domain.[183]

While the demand, demonstrated through the supervenience-argument, that "to cause any property (except at the very bottom level) to be instantiated, you must cause the basal condition from which it arises (either as an emergent or as a resultant)"[184] is believed by Kim to be satisfactory answered through his commitment to the reductive identity of any physically respectable property to the elements treated by physical science, the idea that there can be any kind of reflexive downward causation that arises out of distinct higher-order levels themselves strikes him as incoherent. The type of downward causation that non-reductive theories demands cannot, on his view, be anything but paradoxical. It is not the kind affected by the baseball as a macro-object causing a disturbance in the molecules of the glass it breaks, but downward causation in its reflexive sense, that of higher-order properties exercising causal powers on their own lower-order constituents that Kim considers to be incoherent,

[181] Kim (1998) page 118
[182] Ibid. page 110
[183] Ibid. page 119-120

[184] Kim (1999) page 24

and this incoherency arises from the apparent circularity present in such causal relations;

> After all, higher, level properties arise out of lower level conditions, and without the presence of the latter in suitable configurations, the former could not even be there. So how could these higher-level properties causally influence and alter the conditions from which they arise? Is it coherent to suppose that the presence of X is entirely responsible for the occurrence of Y (so Y's very existence is totally dependent on X) and yet Y somehow manages to exercise a causal influence on X?[185]

In this type of downward causation there is a mutual interdependence between the upward determining forces of the lower-order constituents and the downward causing forces working on these very same constituents, something which Kim terms an apparent absurdity.[186] His argument against such downward causal force of the whole on its parts is structured along the same lines as the supervenience-argument, asking what work, given that the emergent arises from certain basal conditions, is left for the emergent that the lower-level base can not do itself, leaving the emergent causally inert. He says; "If an emergent, M, emerges from basal condition P, why can't P displace M as a cause of any putative effect of M? Why can't P do all the work in explaining why any alleged effect of M occurred?".[187] In his view, postulating any higher-order causes in addition to those operating on the basal level creates an overabundance of causes, and so according to the principles he considers as defining physicalism, the causal status of the dependent event is threatened by that of the event it depends on.[188] As I will attempt to show in the following sections, it is possible to conceive of mental properties playing an indispensable role in the production of behaviour, and it is because without the presence of the mental as a downward force on the basal conditions these lower-order properties cannot do the job required of them. A job that is too great for the properties prescribed causal efficacy on Kim's reductive account. Kim admits that his reductive identification of higher-order properties such as those of a mental character with their underlying realizers does not provide an all-encompassing solution to the mental causation issue or the mind-body problem in general, but he believes that it does better than any suggestion that the

[185] Kim (1999) page 25
[186] Ibid. page 28
[187] Ibid. page 32
[188] Kim (1998) page 53

non-reductionist can offer. As he demonstrates in his supervenience-argument, the non-reductionist need to come up with a credible way of presenting downward causation if they are to avoid falling pray to the charge of epiphenomenalism, or accept that he has come up with the best possible solution to a pervasive problem;

> The position is, as we might say, a slightly defective physicalism- physicalism manqué but not by much. I believe that this is as much physicalism as we can have, and that there is no credible alternative to physicalism as a general worldview. Physicalism is not the whole truth, but it is the truth near enough and near enough should be good enough. [189]

SECTION THREE

3.1 Alicia Juarrero: The self-determination of dynamic systems

Kim's theory of mental causation is fundamented on his view of physicalism as ontological physicalism, where the world is believed to be of nothing but bits of matter and their aggregates, constrained by the laws of physics. Through the demand of reductive identification of any event or object above this level with the components describable in terms of basic particles and the laws governing them he leaves no room for causation that is not acquiescent to the belief that causal relations concern only determinate spatiotemporal pairing-relations between pairs of physically describable substances.[190] This view of causation implies an ontological bias that favours the parts over that of the whole and that gives priority to existents intrinsic particulars over context-related processes. Alicia Juarrero represents a school that challenges these assumptions by claiming that conceptualizing all existents as fundamental particles or aggregates of such does not capture the nature of complex, herein biological, systems. The claim is that such systems demonstrate emergent macroscopic properties that de facto can not be derived from the laws and theories pertaining to the microphysical

[189] Kim (2005) page 174
[190] Ibid. page 86

level.[191] They can only be understood by way of reflexive downward causation of the whole on its parts and this causal power can not be incorporated into Kim's causal scheme. Juarrero and Kim have in common that they both consider causation as generation, or effective production[192] versus mere counterfactual dependence, and they both agree that postulating two separate chains of causal power is a non-starter. But unlike Kim, Juarrero believes that there exists non-decomposable (as in not exclusively microstructurally defined) properties that have truly emergent causal powers in a physically relevant way. That is, the appearance of higher-level causality is more than just "regularities arising out of supervenience on causally linked lower-levels"[193] and higher-order powers[194] can claim a substantial ontological status without tipping over into dualism. Accordingly, causality does not concern just pairing-relations among spatio-temporal particulars, although they are always realized through them. The central point is that such higher-order causal powers play an indispensable role in affecting the behaviour of complex systems, providing them with the conditions for their very existence.

Juarrero synthesises findings in systems theory, nonlinear mathematics and emergence theory in an account of how complex systems as a whole can claim to a degree to be self-causing by way of the systems downwards efficacy on its parts. Unlike Kim's functional or macro-properties that are explainable ultimately in terms of their basal conditions and the fundamental physical laws describing these,[195] the dynamic systems of Juarrero's account are constituted by a kind of circular, inter-level

[191] Juarrero (2009) page 85

[192] Kims understanding is however closely tied to his belief that genuine causation only occurs in the domain describable by physics; it is to be understood as energyflow or momentum transfer. (Kim (2005) page 47)

[193] Kim (2005) page 21

[194] The term properties is problematic in dynamic systems theory as it is a static term, and this theory is fundamented on dynamics or change. In this regard it is problematic to speak of properties as they are atemporal, while emergence is a temporal process (Thompson (2007) page 418). As Thompson states; "Strictly speaking, it does not make sense to say that a property emerges, but only that it comes to be realized, instantiated, or exemplified in a process or entity that emerges over time" (Thompson (2007 page 418). I will however use the term here but I am aware of its inadequacy in describing emergent elements.

[195] Kim (1999) page 20-21

causality, where system-level properties created by interactions among dynamical processes regulate and constrain the behaviour of the lower-level components;

Second, when parts interact to produce wholes, and the resulting distributed wholes in turn affect the behaviour of the parts, inter-level causation is at work. Interactions among dynamical processes can create a systems-level organization with new properties that are not the simple sum of the components that create the higher level.[196]

Whereas any higher-order properties in Kim were nothing but the causal power of its microphysical realizer at a given time, the systems-level organizational powers here can not be understood solely on the basis of the parts making up its material composition. The interactions of these components create unique powers that are not reductively explainable, but they are still physically explainable in the sense that they can not be considered as immaterial.

In an article titled "Top-Down Causation and Autonomy in Complex Systems" Juarrero describes how such causation comes into play in dynamic systems. What characterizes such systems is a combination of flexibility on the part of the components, and the cohesion and robustness of the properties that integrate these parts into a unified whole. They are also characterized by evolvement, a constant movement to a greater degree of complexification, which involves decoupling; the properties of the whole are taken further away from the fundamental energetic forces that shape its basic material components, which implies increased independence from microphysical forces expressed in what physics terms the laws of nature. What then, are the conditions for such movement towards organizational self-determination, pinnacled in the achievement of human rational agency? The fundamental characteristic of systems with this potential is that they are sensitive to conditions other than those specified microstructurally**.** The entities treated by Kim are static entities, devoid of dynamic potential. Being defined exclusively microstructurally as describable in terms of the basic particles that constitute them, the intrinsic properties of these particles and the relations that configure these particles into a macro-

[196] Juarrero (1999) page 5-6 (as quoted in Murphy and Brown (2009) page 86)

structure,[197] such structures are rigid entities that are isolated from interactive influences derived from context and historicity. Juarrero puts it this way;

> An agglomeration or conglomerate composed of particles independent of each other at equilibrium cannot differentiate into a complex organization with emergent properties, and particles related only in terms of relative position at best produce agglomeration or conglomerates. Since the properties of particles do not change when they are merely elements of a conglomerate, any novel characteristics of aggregates near equilibrium- such as temperature and pressure- are merely nominally emergent features of the statistical average of the large number of particles.[198]

Movement is generated by openness, a quality that aggregates do not possess.
Any inputs and outputs from such entities are nothing but elements of causal chains created by the parings of spatial affinity of their physical realizers, fuelled by energetic exchanges of basic particles. The parts making up the entity are not affected by the whole they partake in, and so the system as a whole remains closed off from interacting in a significant way with the environment. This closure means that such entities can not become ordered or complex through this become decoupled from the energetic exchanges from which they arise. It is only through openness that new levels of organizational complexity with novel downward affective power can come into existence. The externally relation-character of such open systems means that they can not be supervenient in Kim's sense[199] and the persistence of such open systems is fundamentally dependent on its relational characteristics, wherein the persistence is "a persistence of an organization of process, not of the constituents that undergo that processes"[200]. In other words, the understanding of components making up the material system is secondary to the understanding of how the system relates to its contextual conditions.

The distinction between open and closed systems was originally proposed by Ludwig von Bertalanffy in the 1950s, the characteristic of an open system being that there is an inflow and outflow and therefore a change of the component materials.[201] Contrary to his belief though, open systems are identified not only in the realm of biology but

[197] Kim (1999) page 6-7
[198] Juarrero(2009) page 84
[199] Andersen et.al (2000) page 333
[200] Ibid.
[201] Skolimowski (1974) page 211

also in that of chemistry. Ilya Prigogine and his associates received the Nobel Price in chemistry almost thirty years later for their studies of far-from-equilibrium dissipative structures.[202] The discovery of how previously uncorrelated particles or process suddenly become coordinated and interconnected[203] suggested an answer to how complex systems come about in the first place. The key to such jumps to higher levels of complexity (i.e. greater degree of interconnectivity) was seen as in the entities being non-linear open systems far from thermodynamic equilibrium. The basic form of downward causation takes place when open systems are driven far from equilibrium as a result of exchanges of matter and energy with their environment.[204] When the system reaches a critical point it will precipitate a phase change[205] through which a top-down restrictive influence on the degrees of freedom of the constituent particles will emerge.[206] The behaviour of the particles is constrained by its participation in a global structure. Being non-linear, the behaviour of such systems can not be predicted nor explained in terms of input-output relations (in the way Kim does) because the value of the input is not directly proportional to the sum of the inputs.[207] Or, stated in another way; "every instance of non-linearity is an instance that cannot be derived aggregately from lower-level aggregates, i.e. every instance of non-linearity is an instance of emergence".[208] This is because the movement generated by the systems openness means that once the system has been started up at a particular point, its trajectory never returns to that point.[209]

Open systems are necessarily those that are inherently far-from-equilibrium and therefore require constant interaction with an environment to be able to exist over time. Attempting to describe these with the explanatory resources of Kim's reductive approach unavoidably leaves something important out, because the emergent macro-properties of such systems cannot be derived from the laws and theories pertaining to

[202] Dissipative systems are chemical systems where a sudden change from random motion to order emerges as the system is brought far from equilibrium (Juarrero page 84)
[203] Murphy and Brown (2009) page 85
[204] Juarrero (2009) page 84
[205] A phase change is the sudden shift in into a new mode of organization as a result of beign driven far from equilibrium (Juarrero 84)
[206] Juarrero (2009) page 84
[207] Thompson (2007) page 39
[208] Andersen et.al (2000) page 343
[209] Walter (2009) page 168

the micro-physical level.[210] [211] There is a substantial difference between the weakly emergent systems at the chemical level and the complex biological systems that actively modify their environment in order to ensure their continuous self-construction and persistence.[212] The requirement of openness is however common to all such complex systems. There must be an element of interaction with the context the system is embedded within, where the system finds its conditions for self-preservation and those elements that create and sustain it, and actively selects these elements.[213] The top-down control of the these interactions increase as one goes up the organizational ladder, but this complexification is at bottom fuelled by open systems being driven far from equilibrium as a result of exchanges of matter and energy with their environment. Increased top-down feedback from the system to its constituents involves increased decoupling from energetic constraints on this interaction, and allows the system more autonomy to select the elements in its environment that it is to be affected by, but the downward causation of such complex dynamic systems is never severed completely from that which can be included within an broad version of physicalism.

3.2 top-down causation as context-sensitive constraints

The type of causation that comes into play as previously independent elements are coupled and integrated into a complex whole is not the effective causation of determinate microstructural entities participating in a causal chain, but rather conceived of in terms of constraints. Greater complexity is achieved through a greater degree of control of the constituent elements participation in the whole, where the selective criteria that emerge from the interactions of the system with its context exercise a restraining influence on the movements or freedom of its components. The emergence of context-sensitive constraints "take the system away from independence by making the elements comprising the system interact in such a way that their behaviour depends on one another's- and on what went before and what is occurring

[210] Juarrero (2009) page 85
[211] As stated previously physics is traditionally concerned with systems that are described as closed.
[212] Juarrero (2009) page 94
[213] Ibid.

around them in the environment.".[214] The causal property of the system to effect the interaction of its parts is not the power of some overriding entity to reach into an isolated and causally closed system but is a product of the systems feature of openness and the context it is embedded within. It is defined by the conditional probabilities that that describe the range and behaviour of microstate arrays; "The novel complex interaction just is the changed probability distribution of the components' state space".[215] These constraints retain the systems integrity by restraining freedom of its microphysical components without eliminating it, and by this maintain the global dynamics of the system to an optimal degree (which means more freedom for the system as a whole). This causal power is genuinely emergent in that it is not reducible to an aggregation of the lower-level constituents causal power, for as Juarrero states; "The causal relationships that the new codes specify about the higher level are for the most part sealed of from the energetic-type causes operating at the lower level" [216], and so the higher-level patterns that come about as a result of the interchanges of the system and its environment is not reducible to the mass effect of lower-level constituents.

Since this kind of causation is not definable within the parameters of Kim's ontological physicalism, what is to be added in order to describe such higher-level downward causation? As mentioned it is the context and historicity of a system that affects its organizational, or higher-level patterns**,** in relation to its material constituents. Constraints are constructed both by the open systems' embeddedness within a larger context[217] and on what has gone before (the systems initial conditions), represented by the changed probability-distribution of the microphysical components.[218] These kind of context-dependent constraints are often described in terms of ontogenic landscapes, where systems are visualized independently of their substance and spatio-temporal scale and components described in relational terms. Here we speak of topological rather than substantive entities. It is important to understand that such organizational constraints are independent of specific physical

[214] Juarrero (2009) page 86
[215] Ibid.
[216] Murphy and Brown (2009) page 53

[217] Context shoud be understood in terms of encompassing systems**.**
[218] Juarrero (2009) page 87

configurations, although they are always realized in the physical space.[219] The requirement that organizational constraints are always realized in physical systems means that physicalism as such is not abandoned but it would entail giving up on the "completeness of physics"[220] principle which informs Kim's conception of supervenience and his demand of decomposability. So physicalism is not here exclusively related to the principles expressed by microphysics, but it is still concerned only with the goings-on of the physical world.

3.3 Ontogenic landscapes and top-down feedback

An ontogenic landscape is a systems phase space, an abstract space whose coordinates are the degrees of freedom of a system's behaviour. A system near equilibrium would have a completely flat landscape, representing an object with no propensities or dispositions. Such dispositions are expressed as attractors, which are defined as a set of points in a phase space in which the trajectory of the system flows towards.[221] An open dynamic system made up of configural regularities affecting constituent interactions develops quite a different topography from that of a closed near-equilibrium system. Juarrero describes how such a phase space appears, in the following quote:

> In contrast, the increased probability that a real system will occupy a particular state can be represented as wells (dips or valleys in the landscape) that embody attractor states and behaviours that the system is more likely to occupy. The deeper the valley, the greater the propensity of its being visited... Topologically, ridges separating basins of attraction are called *separatices* or *repellers.* Sharp peaks are called *saddle points* representing states and behaviours from which the system shies away and in all likelihood will not access; the probability of their occurrence is lowered or eliminated altogether... Separatrix height represents the unlikelihood that the system will switch to another attractor given its history, current dynamics, and the environment. The steeper the separatix's walls, the greater the improbability of the system's making the transition. On the other hand, the deeper the valley, the stronger the attractor's pull, and so the more entrenched the behaviour described by that attractor and

[219] This non-identity of the properties defining the relational pattern of the global complex structure with the properties of its material basis of course implies that there is no problem of multiple realizability here. (Juarrero (2009) page 89)
[220] "all microphysical events are determined by prior microphysical events and the laws of physics" (Juarrero (2009) page 91)
[221] Walter (2009) page 168

the stronger the perturbations needed to dislodge the system from that propensity. The broader the floor of a basin of attraction is, the greater is the variability in states and behaviours that the attractor allows under its control. The narrower the valley is, the more specific is the attractor, that is, the fewer state and behaviours within its basin.[222]

This paints quite a different picture of than the mechanic instance-causation of Kim, where any higher-level causation is made out to be mereological aggregates of linear relations among its intrinsic constituents. The details of the systems material components come second to its dynamics of process, and the causality affecting these dynamics is a negative or constraintive power. Unlike the lower-level energetic type causes determined by the laws of physics, such system-level forces are understood as restraints reducing the possibilities of the pathways that the system may follow. There are no forces from above operating here, no "vital entelechies", just feedback from the features of the whole system to the architecture of its components. This type of causation is circular, in that it involves an interdependence of the components processes and the organizational unity that they comprise, but avoids the problems marring downward causation viewed within Kimean supervenience-perspective. **C**omplex adaptive systems are characterized by recursive feedback-processes in which the product of the process is necessary for the process itself, precisely the kind of circular causality that Kim considers incoherent. On Juarrero's account this kind of causality is rather a genuine organizational force, operating not as the efficient causes treated by classical physics but in terms of meaningful criteria determined at the higher level.[223] The microphysical configuration of the network exercises its causal efficacy and produces a particular output in virtue of being entrained into a higher-level dynamics that thereby embody emergent semantic features,[224] where the semantic content of such features is integrally related to its contextual embeddedness and external interactions; "for the ideas of *feedback, information, environment, past experience* are normative categories; we cannot make sense of them in any purely physical system, but only in a system that admits values, evaluations and norms"[225]. We see here the introduction of semantic and normative aspects of higher-order constraints, describing the parameters of the global, higher-level criteria for such

[222] Juarrero (1999) page 156 (as quoted in Murphy and Brown (2009) page 76-77)

[223] Juarrero (2009) page 88

[224] Ibid. page 90

[225] Skolimowski (1974) page 221

constraints. Even in simple recurrent networks such as artificial word-reading systems the context-dependence of grammatical constraints operates as feedback-loops constraining the output of the system by way of semantic or informational criteria. When such a word-reading system is trained to learn the frequency of occurrence of each possible successor word for each possible context, it reveals the development of an internal dynamic that reflects grammatical category and meaning.[226] The output of the network is in other words constrained by emergent semantic features, these setting criteria of suitability that determined the output. Even in such a simple system the output is not determined by intrinsic variable configurations but the information such configurations carry when integrated into a context-sensitive system.

The informational characteristics that determine the constraintive direction of the higher-order organizatory power differs on account of the degree of internalization of such top-down constraints. In weakly emergent systems such boundary-constraints are imposed exogenously, as in the above-mentioned word-reading network, whereas genuinely emergent systems such as the living organism determines the criteria of suitability guiding top-down selection endogenously. The regulatory processes that constrain the direction of its components are here produced by the systems own dynamics rather than its values being set by external influences. They exhibit a degree of autonomy not seen in the weakly emergent systems, this being based on their ability to "create the very constraints that control the matter-energy flows that make the structure possible".[227] On Juarrero's account the key notions are here are fitness and normativity;

> Instead of processes being determined by energetic considerations alone, selection based on criteria of suitability determined at the higher level defines a direction that is increasingly autonomous and decoupled from merely energetic considerations.(…). The dynamics themselves also select and delete components according to fitness criteria determined at (and meaningful for) the level of the coherent whole.[228]

[226] Beer (2000) page 91
[227] Juarrero (2009) page 92
[228] Ibid. page 95

In this "formal cause-like process"[229] of top-down selection a normative process is at work. The notion of normativity or fitness is intimately connected with that of movement or activity. In self-organizing systems movement is always directed by the goal of "maintaining stable patterns despite constant change in the matter in which they are composed and despite perturbations",[230] and so any change is steered by the normative criteria of self-preservation or maintenance, in relation to the context the system is embedded within. The fitness-criteria is the persistence of the whole,[231] and so the restraining of lower-order components and boundary conditions is done on the basis of such normative criteria.

This internalization of regulatory processes is a gradual evolvement, starting with the emergence of chemistry, where the selecting of component molecules in such a way that the far-from-equilibrium conditions necessary for its dynamical persistence is maintained is the minimal sense of such normative function. The internalization of the criteria of suitability that top-down selection of lower-level components is carried out according to gives the system more freedom in determining the contextual interactions it can enter into.[232] Systems where the constraints that control the matter-energy flow that make the structure possible have been fully imported into the system are termed autopoietic. The criteria of top-down control and the related robustness in maintaining its organizational identity in the face of external perturbations must here be understood in relation to how the system is structurally coupled with the context it is embedded within. Intentionality and normativity is here defined in terms of the 'goal-directed' process of self-maintenance of organizational or operational closeness within a domain of interaction specified by these very fitness-criteria. Thus emerges "the strong autonomy present in biological hereditary autonomous systems".[233] The mechanism of this relation is the subject of the next section.

We seem to have here a prime candidate for the kind of casual power that we demand of that which is to be considered in connection to agency and mental causation, but it is still a long way of that of the cognitive powers of the human mind. Juarrero's

[229] Juarrero (2009) page 95
[230] Murphy and Brown (2009) page 71
[231] Juarrero (2009) page 93
[232] This point will be explained more thoroughly below
[233] Juarrero (2009) page 93

account does however suggest the fundamental prerequisites of such cognitive capacitates and we can imagine the continued decoupling of complex systems from energetic forces, and the adjacent emergence of informational or normative mechanisms could end up with rational biological creatures such as us. Before I explore the ideas of dynamic openness and organizational closeness in relation to cognitive capacities of sentient beings I would like to present the origins of the idea of organisms as autopoietic structures of process, and introduce the idea of cognition as the relationary process of such systems to its context.

3.4 Autopoiesis- the organization of the living

The term autopoiesis was introduced by the Chilean neurobiologists Humberto Maturana and Francisco Varela in a paper written in 1974. They were concerned with disclosing the nature of the living organism, by way of revealing "the organization of living systems in relation to their unitary character",[234] and looking at how cognition is a biological phenomena constituting the realization of such a unity in the physical space. Rather than considering the physical structure of biological systems and the properties of these components they looked to how the spatial relations between structural components are specified by the network of processes of production of these components.[235] The term autopoiesis was coined in this connection in order to capture the central features of the living, namely autonomy and self-production,[236] the two characteristics of its organization. Their goal of divorcing the term organization from any mystical or transcendental connotations and instead using it to refer to the specific relations that pertain to an autopoietic system is reflected in Juarrero's effort to deny the evolvement of forces over and above the "processes interlaced in the specific form of a network of productions of components which realizing the network that produced them constitute it as a unity."[237] The claim that it is the organization of living systems, rather than the physical constituents realizing this network that "holds

[234] Maturana and Varela (1980) page 75
[235] Ibid. page 80

[236] poiesis means creation or production
[237] Maturana and Varela (1980) page 80

it together" (as in gives it its identity) is in Maturana and Varela's work explained through the use of the term machine. By this they remove themselves from any vitalistic implications. They start of by comparing living systems to artificial machines. In any machine the organization is independent of its structure, that being the components which integrate a concrete machine in a given space. The organization does not first and foremost specify the properties of these components, but rather the relations which these must generate to constitute the machine as unity.[238] Unlike an artificial or man-made machine however, living systems have as an essential characteristic that the product of their organizational function is not something different from themselves. Their goal is self- maintenance, not the production of some external result. The goal-states of artificial, or allopoietic machines are defined by something external to it and so it has no autonomy because its organization is subservient to conditions determined externally, whereas the living system set the variables in accordance with its own demands for self-production.

According to Maturana and Varela, the living system maintains its organizational integrity by a continuous production of components that are involved in the transactions and transformations which constitute the system.[239] The idea that the identity of the system is preserved on the organizational level rather than the level of physical structure mirrors Juarrero's explication of how it is the downward forces of this integrity that holds the system together. Maturana terms such stability-maintaining systems homeostatic machines, defined as a device for holding a critical systemic variable within physiological limits,[240] the critical variable here being the systems own organization:

> An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produce the components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network".[241]

[238] Maturana and Varela (1980) page 77
[239] Ibid.
[240] Ibid. page 66
[241] Ibid. page 79

Just as in Juarrero's account the dynamic or active nature of this network of processes of production is emphasised rather than the focus being on the instantiations of properties in entities tied to determinate spatio-temporal positions, as was the basis of Kim's supervenience-view of higher-order causality;

> (…) autopoietic machines are unities whose organization is defined by a particular network of processes (relations) of production of components, the autopoietic network, not by the components themselves *or* their static relations. Since the relations of production of components are given only as processes, if the processes stop, the relations of production vanish; as a result, for a machine to be autopoietic, its defining relations of production must be continuously regenerated by the components which they produce.[242]

Although we can divide physical autopoietic machines into parts this does not mean that we can explain the operations of it on account of these parts as it "does not reveal the nature of the domain of interactions that they define as concrete entities operating in the physical universe".[243] In Kim's account of higher-order systems on the other hand, the demand of reductive identity with their microstructural realizers means that they are sufficiently described by the parts and their interactions. Here the hierarchy is turned on its head as material particulars are considered as parts through their role as components realizing the systems organizational unity.[244] The consistency of the autopoietic machine that Maturana and Varela stipulate is not on determined by its maintaining its components constant (being static), but by maintaining constant certain relations between components otherwise in continuous flow or change, that is, maintaining constant the relations that define them as autopoietic.[245] Living systems differentiate themselves from the non-living by being autonomous, in the sense that the product of their functioning is the maintenance of their organization, unlike allopoietic machines that have as the product of their functioning something different from themselves. This means that the conditions of such maintenance are not determined exogenously, but by the organizational nature itself. The capacity of

[242] Maturana and Varela (1980) page 79
[243] Maturana and Varela (1980) page 82
[244] "The spatial domain of relations is logically secondary to the domain of relations between the processes of production of components*:* spatial relations between the components are specified by the network of processes of production of components which constitute its organization. (Maturana and Varela (1980) page 80)
[245] Ibid. page 81

autopoietic machines to specify their own boundaries by way of their operations of self-productions makes them unities in a way that allopoietic machines are not. Their boundaries are specified by the observer, who in specifying its output and input surfaces specifies what pertains to it in its operations.

The elements of autonomy and self-determining identity must be understood in relation to Maturana and Varela's use of the concept of cognition, which they consider to be the defining characteristic of the implementation of an autopoietic system as a physical system; "autopoiesis in the physical space as a necessary and sufficient condition for a system to be a living one"[246] Maturana says, implying that the process of autopoietic self-production must always take place in a physically instantiated system. So even though organization is logically prior to structure in the autopoietic understanding of the living, such a system must always be realized as a physical structure in order to be characterized as living. The notion of physical space in the above quote does not however refer exclusively to the internal structural composition of the system, but just as much to the physical space which it must interact with in order to preserve its organizational unity or identity.[247] The organizational closeness that provides the autopoietic system with its integrity is interdependent with a material openness that subserves the maintenance of homeostasis. Cognition is the relation that binds the two together.

Maturana says that; "living systems are cognitive systems and living as a process is a process of cognition."[248] An autopoietic system as a homeostatic system has as its function to produce and maintain its circular organization by determining that the components which specify it should be those whose synthesis or maintenance it secures[249] and it defines its domain of interactions according to the maintenance of this circularity. The domain of interactions that a living system can enter into is called its cognitive domain and is on Maturana and Varela's account different from its environment, which lies exclusively in the domain of the observer and is defined by

[246] Maturana and Varela (1980) page 84
[247] This domain of interactions is in fact viewed as a lager system, accordingly the boundaries of a system do not necessarily coincide with that of their material composition.
[248] Ibid. page 13
[249] Ibid. page 9

the classes of interactions into which the observer can enter into and which he treats as a context for his interactions with the observed organism.[250]

By considering cognition to be intrinsic to any living system the theory of autopoiesis takes some of the mystery out of the mental. Cognition is here related to a domain of interactions specified by the demands of the living systems organization. It's a far cry from the idea of cognition as the disembodied manipulations of internal representations as is suggested by the view of the mind as 'inner'. The process of cognition is in the autopoietic perspective a fundamental biological function that is necessarily subordinated to the maintenance of homeostatic equilibrium[251]. Because the power to interact with its context is an essential element of the autopoietic system as an open system it is present at all levels of that which can be characterized as living. On this account there is no sudden appearance of cognitive power as one moves up the scale of biological complexity. The notion of cognition as utilized here is however without the characteristics which I earlier presented as defining of the mental. Here the concepts of normativity or purpose belong to the domain of domain of descriptions, reflecting the operations of an observer considering the system in some encompassing context. Unlike Juarrero, Maturana believes that the notion of function in the telenomic sense has no explanatory value concerning the operations of the autopoietic organization. "Living systems" Maturana states, "as physical autopoietic machines, are purposeless systems".[252] According to him, the relations implied in the notion of function are not constitutive of the organization of autopoietic systems but are descriptive notions used to characterize a system in relation to some context of use defined by an observer.[253] What then happens to the idea Juarrero expresses, that elements present in the systems cognitive domain are imported into its structure through a selection process where the fitness criteria is the persistence of the system (as a normative process)? What about the systems apparently goal-directed activity of pursuing selective exchanges within its domain of interactions in the process of maintaining its organizational integrity? It would seem as if this normative selective property is an integral part of the autonomy of any living system, as it must

[250] Maturana and Varela (1980) page 10
[251] Ibid. page 49
[252] Ibid. page 86
[253] Thompson (2007) page 141

be if we are to make sense of mental causation as intentional or semantically guided causation.

The original theory of autopoiesis was meant to apply first and foremost to existents on the cellular or molecular level and so it can be argued that there is here no need for graded norms and goals. It provides us with the all-or nothing norm of self-continuance, but not the graded norm implied by an organism actively seeking to improve its conditions of self-production.[254] It does seem hard not to recognize an aspect of adaptivity or function in the self-maintaining activity of living systems at levels above the bottom, a richer sense of agency than that implied by the a minimal purpose of continued self-maintenance. As I will relate further on, Valera later came to believe that normative notions are constitutive properties the whole system possesses because of the way the system is organized, and so that it is not simply descriptive but also explanatory.[255] This requires more than the "conservation of identity through internal material turnover and external perturbations to the system".[256] In addition a minimal agency, in the sense of the capacity of the organism to actively improve its conditions of self-production must be present, a capacity that all biological beings share.

3.5 Self-organization and top down causation in biology

Those that consider cognition or the mental in a naturalistic perspective want the understanding of this term to refer to the biological domain, in the sense that it is in this area that many of the explanatory resources for making cognition intelligible is taken from. In the words of the cognitive scientist Randall D. Beer; "if it deals with a living system it must be about biology."[257] This entails reduction to a certain degree but not causal reductionism as implied in Kim's account. Its ontological reductionism as adherence to the idea that there are no "non-physical" forces at play, but also that there are levels or orders above that covered by physics that have genuine causal force

[254] Thompson (2007) page 147
[255] Ibid. page 146
[256] Ibid. page 147
[257] Beer (1980) page 65 (preface to Maturana and Varela)

of a downward kind, and that should be subject to scientific research in its own right. Within biological science there has been an increasing recognition that it is not sufficient to understand biological entities in terms of their physical or chemical composition but that it must be supplemented by the recognition that living beings, whether they be bacteria or rational human beings, share the characteristic of dynamic openness**.** The move from focusing solely on the physiochemical requirements for life, as dictated by the analytical methods of molecular biology,[258] to recognizing the characteristics of the living entity as a system emerging on higher levels of organization reflects the paradigm-shift which was taken up in the section on Juarrero.

It involves an increased recognition that the previously taken or granted initial and boundary-conditions of living systems must be treated as subjects of study in addition to the material conditions that occur within these, as well as a shift from considering complex wholes as aggregates of their parts towards putting emphasis on relational processes over material components.[259] It is a move away from the constraints of the atomist scheme of strict physicalism, and a move towards the claim of non-reductionistic physicalists that the study of biological processes at the level of underlying physicochemical components should be complemented by the recognition that not all processes making up complex systems such as living organisms can be adequately described in physiochemical terms.[260] The reduction of biological processes to their underlying physical and chemical components involves a denial of the existence of cognition as the characteristic of emergent higher levels of organisation[261] because this notion can not be formulated in physiochemical terms. The belief among those that consider teleonomy, in the sense of the normative character of cognition, to be intrinsic to living organisms is that such a reductionistic approach leaves out a central feature of life;

The mechanistic approach has been most successful towards making new discoveries but is insufficient to account fully for biological phenomena. The laws of physics alone can not explain biological processes above the molecular level of organization. There is no need to introduce “vital forces” or any

[258] Ayala (1974) page i
[259] Ibid.
[260] To put it nomologically; not all biological laws can be derived from the laws governing physics and chemistry. (Ayala (1974) page xi)
[261] Ayala (1974) page xiii

other metaphysical entities, but principles of explanation not reducible to those of physics and chemistry are required.[262]

It is particularly evolutionary biology which has taken up these considerations. Fundamented by the Darwinian recognition of the importance of context through the role of the environment, this branch of biology seeks to divorce itself from the restrictions of strict physicalism. By understanding the causal histories of biological structures in terms of the goals of survival and reproduction, evolutionary biology decouples the macro-properties of the living system from their underlying atomic or molecular constituents. Whereas molecular biology claims that the existence and structure of a given part of the organism could only be understood in terms of the micro-processes by which it is constructed in the individual organism, bringing function and goal-directedness into the picture allowed for understanding causal processes on higher levels. The restraintive forces of natural selection give direction to the evolutionary processes by increasing the adaptivness of the living organism to its environment.[263] The term restraint brings us back to Juarrero's elucidation of how higher-order causal powers constrain the freedom of their own components through the limitation of options, thereby giving more freedom to the system as a whole. Biology brings the ideas of genuine higher-order causal powers with reflexive downward efficacy to bear on real-life phenomenon, thereby presenting a challenge to the idea that only micro-particulars dealt with by physics have genuine causal standing. The evolutionary biologist Henryk Skolimowksi expresses the challenge to the reign of physics as universal paradigm as this;

(…) the particular difficulties in which we now find ourselves in the realm of biology, and also in relation to the whole heritage of our scientific knowledge, stem from the restrictive harness of a rationality which is no longer adequate for the recent extensions of our knowledge and for the cognitive needs of contemporary man. The rationality developed under the auspices of physical science is a harness, for it ties us down to a certain conceptual framework and obliges us to observe criteria of validity as accepted by the current scientific rationality, and yet often explain phenomena with more illumination than would have been possible within any physicalist model.[264]

[262] Ayala (1974) page xi
[263] Ibid. page xv
[264] Skolimowski (1974) page 206

He presents a number of attitudes characteristic of the 'old' physicalist paradigm, which reflects those that were presented in the section on reduction, and counters these with an equal number of tenets that express the paradigm within which the studies of the living is increasingly being conducted within. Many of these reflect the principles expressed in Juarrero's account of higher-order causal power, such as; "the methods of physical science are insufficient for the study of the phenomena of life on the high level of complexity", "that we cannot limit ourselves to simple and relatively isolated systems, for life systems are enormously complex and intricately connected", and "that the behaviour and action of many large and complex systems is often inexplicable by the behaviour of the constituents of the system; the total behaviour often equals more than the sum of its parts or differs from this sum."[265].

An illustrative example of research conducted within this non-reductive perspective is Donald Campbell's presentation of the anatomy of the jaws of worker termites.[266] It also serves as an excellent demonstration of Juarrero's whole-part constraint applied to a real-life dynamic system. Campbell was one of the first to use the term downward causation in a scientific context to denote the influence of global, collective-variable dynamics on local behaviour.[267] His account lays out empirically how a larger system of causal factors can exert causal efficacy on lower-level entities by means of selection. The larger system is the organism in interaction with its environment, where feedback from the environment exercises a selective effect on the bottom-up production of genetic variation.[268] Here the potential causal powers of the constituents making up the organization in the physical space are activated by the patterns of organization of these very same constituents. In other words, there is recursive causality at work in a real-life system.

The selective mechanisms at work in his example are the optimising forces that cause the jaws of worker termites or ants to be developed so that they gain the maximum

---

[265] Ibid. page 208 (all quotes from here)
[266] Campbell (1974)
[267] Thompson (2007) page 424
[268] Murphy and Brown (2009) page 65

usefulness in the organism (as well as for the greater system, that of the social organism). The anatomy of the jaws of worker termite functions in accordance with Archimedes' laws of levers, optimising the form for the survival of the organism,[269] while the formation of the particular proteins of the muscle and shell of which the system is constructed is determined by molecular and atomic coupling processes.[270] The higher-order law of levers, operating as part of the complex selective systems at the level of the organism, is needed to explain the particular distribution of proteins found in the jaw and hence the DNA-templates guiding their production. All the while the implementation of such higher-level laws is wholly dependent of the bottom-up production of genetic variation and in this case the protein structures it produces. So in order to understand why such features are developed it is not sufficient to consider the micro-constituents that make up the building blocks of such structures, but also the function it serves in the larger system.[271] The downward causation occurs through the restraints imposed by the selective system of the highest level of selection that affect distributions on all lower levels.[272] In this case it is a selective system operating to sustain a specific level of organisation that constrains the protein distribution in social-insect jaws[273]. We can see here a normative process at work, in the sense that "(…) components are chosen on account of their fitness regarding organizational unity".[274]

Campbell considers himself an ontological reductionist, stating that the higher-order organizing characteristics of systems can be explained by natural processes,[275] while rejecting the belief that such higher-order causal powers can be explained on account of the laws governing its physicochemical rudiments.[276] This entails that he accepts that all higher levels act in conformity to the laws of lower levels and the teleonomic achievements at higher levels require for their implementation specific lower level

[269] This is not necessarily straightforward, given that it often entails that worker-ants cannot feed themselves and therefore are dependent on others for food, the foodgivers in turn being dependent on the existence of worker-ants. This require laws of sociology for their explaination. The understanding of context as a 'higher-order' system encompassing the lower that it is embedded within it is here illustrated.
[270] Campbell (1974) page 181
[271] Murphy and Brown (2009) page 57
[272] Campbell (1974) page 182
[273] Ibid. page 183
[274] Juarrero (2009) page 95
[275] Ayala (1974) page xiii
[276] Campbell (1974) page 179

mechanisms and processes which must be specified before explanations are complete. However, an additional principle applies to biological systems in which the constraintive force of selection is involved. This is the emergentist principle, which states that biological evolution encounters laws operating as selective systems which are not described by physics and inorganic chemistry and that show downward causal efficacy;

> Where natural selection operates through life and death at a higher level of organisation, the laws of the higher-level selective system determine in part the distribution of lower-level events and substances. Description of an intermediate-level phenomenon is not completed by describing its possibility and implementation in lower-level terms. Its presence, prevalence or distribution (all needed for a complete explanation of biological phenomena) will often require reference to laws at a higher level of organisation as well.[277]

In other words, scientific description is still incomplete when all the details of the micro-mechanisms involved have been figured out[278] because when we understand the function of these organisms in terms of complexity and hierarchy we must go beyond their physicochemical rudiments.[279] According to the evolutionary biologist, the teleonomic character of living beings, that is the adaptive nature of organisms and their features,[280] resides fundamentally in the goals of survival and reproduction[281] and so when we speak of goals it is for the most part not the conscious, self-reflective value-setting of human beings we are concerned with. Rather it is the normative or semantic content of the cognitive relation that obtains between the organism and its domain of interaction. The science of biology gives empirical content to the ideas of Juarrero and Maturana by utilizing them as explanatory tools in understanding the mechanisms of life. Introducing normative concepts allows biologists to describe without mystification living entities guided by specific values and directed to specific goals[282] and so provides the fundamentals for understanding how semantic or normative factors can play a causal role in directing behaviour in organisms at different levels of complexity. The intentional or cognitive capacities characteristic of

[277] Campbell (1974) page 180
[278] Campbell (1974) page 182
[279] Skolimowski (1974) page 212
[280] Ayala (1974) page x
[281] Murphy and Brown (2009) page 71
[282] Skolimowski (1974) page 219

sentient beings is fundamentally structured around the goals of self-maintenance and reproduction, where selection plays the constraining role of optimizing the functions of organisms through determining the distribution of lower-level events and substances.[283] The notion of selection will not be treated further here, but as I proceed to bring the capacities of cognition and intentionality towards the human realm in the next section, it is implicit as the underlying mechanism fuelling these evolutionary developments.

SECTION FOUR

4.1 Mind in life- making the mental matter

The previous sections have explicated the fundamental requirement of openness and the autonomy or capacity for downward control that emerges from it, as well as cognition in its minimal sense as that defining the relation between the living entity and its context of operations. In this section the goal will be to synthesise these two notions so as to gain an understanding of how cognitive capacities can have legitimate causal power. In order to achieve this I will present Evan Thompson's idea of how mentality emerges in self-organizing dynamic systems by way of the circular causality of continuous sensori-motor interactions involving brain, body and environment.[284] His claim is that the mental can not be understood simply as brain-processes inside the head but must be fundamented on the cognitive notions implicit in the autopoietic organization of biological life and how this finds sentient expression in the self-organizing dynamics of perception and action;[285]

[283] Murphy and Brown (2009) page 58
[284] Thompson (2007) page 10-11
[285] Thompson (2007) page ix

Life and mind share a set of basic organizational properties, and the organizational properties distinctive of mind are an enriched version of those fundamental to life. Mind is life-like and life is mind-like.[286]

The theory to which Thompson subscribes is called embodied dynamicism[287] and arose in the 1990s as a reaction to earlier approaches to cognitive science. Previous models of the operations of mind had portrayed cognition as disembodied and abstract, focusing on computational representation in the mind-brain.[288] The earliest of these modelled cognition as a computer, cognitive or mental processes being viewed as the manipulation of symbolic representations in the brain. According to this approach sensory inputs are mapped onto symbolic representations and these representations are then manipulated in a purely formal or syntactic fashion in order to arrive at a output or solution to a relevant problem.[289] Although it was an improvement from its predecessor, behaviourism, in that it at least recognized brain-states as elements of interest, this perspective of cognition as serial information processing by way of the discrete causal effectiveness of local unit activity in linear processes of rule-governed transformation of static structure into another[290] turned out not to reflect the observations of the brain in operation. In addition to lacking empirical support it created as well an explanatory gap between sub-personal, computational cognition and conscious or subjective mental states that demanded explanation.[291]

Unsatisfied with the approach of explaining neural operations by way of the physical-symbol model, cognitive science turned to connectionism in the early 1980s and it has since remained the dominant approach to the exploration of mental processes.[292] The focus is here on the architecture of neural networks, characterized in terms of units, layers and connections, and the learning rules and distributed sub-symbolic

[286] Thompson (2007) page 128

[287] It is also refered to her as the enactive approach
[288] Ibid. page 10
[289] Ibid. page 5

[290] Walter (2009) page 121
[291] Mental processes are here only concerned with the formal or syntactic properties of mental representations, not with their semantic properties. (Walter (2009) page 119)

[292] The main disagreement is over the nature of computation and representation- symbolic for cognitivism, subsymbolic for connectionism (Thompson (2007) page 10)

representations that emerge for the networks that are activated.[293] The networks are mainly modelled as feed-forward, flowing from an input layer to an output layer, where information processing occurs by activity expanding through distributed networks of neurons.[294] An illustrative example of the perspective of cognitive science is taken from a psychology text-book, where cognition is expressed as mental computation;

> A central idea behind cognitive science is that he human cognitive system can be understood as though it were a giant computer engaged in complex calculation. Just as a computer's complex calculation can be broken down into a set of simpler computations, such as storing, retrieving, and comparing symbols or representations, so a person's action can be broken down into a set of elementary mental components.[295].

This perspective on cognition is reflected in Kim's view that the mind is reducible to "patterns of electrical activity in some group of neurons"[296] where mental properties are specified in terms of their roles as causal intermediaries between sensory inputs and behavioural outputs.[297] A greater recognition of the interconnectivity of neuron-populations filling the roles of the mental components is offered by connectionism than the previous model of information-processing. From the perspective of embodied dynamicism however, the connectionist model does not fully capture the complexity of the connectivity of components making up the neural architecture. As well, by appealing to information-bearing states inside the system as internal structures that encode context-independent information about the world, cognition being the processing of such information,[298] many issues concerning intentionality, language and mind-world relations remain. The main point of contention between connectionist models of mind and dynamic approaches is that, according to the latter, the objectivist notion of information is an inaccurate way of understanding how these representations of the world come about. Rather Thompson takes information to be essentially relational, saying that "information is the intentional relation of the system

[293] Thompson (2007) page 9
[294] Walter (2009) page 115
[295] Atkinson et.al (2000) page 17
[296] Kim (2005) page 71
[297] Kim (1998) page 19
[298] Thompson (2007) page 52

to its milieu".[299] Here information or the semantic aspect of things are not given to a disembodied cognizer and mapped onto representational networks as assumed by computational models of mind. As a reaction to these representationalist models of mind the embodied dynamical model posit mind as an essentially embedded system in the world, extended through the body and looping through the material, social and cultural environment, rather than as the implementation of input-output directed information-processing capacities in the brain. The intentional relations by which the living system interacts with its milieu is established on the basis of the systems autonomy (its organizational closure), and it maintains this organizational identity by acting in accordance with how the elements of its cognitive domain are semantically defined by the demands of this autonomy.[300] In other words, a cognizing being "(…) does not process information in a context-independent sense. Rather, it brings forth or enacts meaning in structural coupling with its environment".[301] The operations of the mind are here not pictured as that of mapping in input at one time onto an output at a later time, but always to maintain appropriate change.

This approach claims to dissolve some of the tension between the properties exhibited by the mind and those exhibited by the material body, a tension that is expressed in Kim's formulation of supervenience as well as in the contention that the properties of mind and those of the body seem to be so different that it is difficult to see how they their natures could be described using the same explanatory resources. While the semantic property instantiated by an organism's intentional state is essentially a relational fact that involves its relationship to external environmental and historical factors, the causal factors conceived as producing behaviour must on Kim's view be intrinsic to the organism. Only the syntactic properties of properties of mental states can be allowed causal power if the requirements of supervenience are to be followed.[302] Herein lays the crux of the problem of mental causation.[303]

---

[299] Ibid. page 59
[300] Ibid.
[301] Ibid. page 58

[302] Kim (1998) page 37-39
[303] "However, semantic properties of internal states are not in general supervenient on their synchronous internal properties, for as a rule they involve facts about the organism's history and ecological conditions. Thus two organisms whose total states at a given time have identical intrinsic properties can differ in respect of the semantical properties they instantiate (…) but prima facie these semantical differences should make no difference to behaviour output" (Kim (1998) page 36)

Considering the operations of mind as essentially interdependent with the activity of the living body means that there is no great mystery of mental causation, the capacity for such downward affective forces being present as a necessary factor from the very beginnings of life. Intentionality, meaning-creation and normativity are parts of the activity of the living being because they as autopoietic systems necessarily must stand in such relations to the domain where their activity takes place. This general idea is expressed in the central tenets of this position, namely that living beings are actively self-generating or maintaining in an autopoietic sense and that this self-maintenance is conducted through the exercise of cognition by way of embodied action. In higher-order organisms such as humans this is done through the nervous system, where re-entrant networks of interacting neurons generates coherent and meaningful patterns of activity, in concert with the recurrent sensori-motor patterns of perception and action. This coupling between organism and environment modulate the formation of dynamic patterns of neural activity, which in turn affects the sensory-motor coupling.[304]

Thompson utilizes the idea expressed by Maturana and Varela that the living system defines a domain of interaction through its organization, to which it relates according to the demands of sustaining viability, the actual acting in this domain being the process of cognition.[305] However cognition in Thompson's account is furnished with more content because he considers this dynamic interaction to involve the creation of meaning. Something becomes informational or gains semantic content through this activity of interaction (structural coupling in Juarreros terms). In Maturana's account there is no 'graded information' being relayed in interactions of the system and its milieu, only perturbations caused by the structural coupling of the system and its environment. The informational content such perturbations have to the system depends on the norms of the systems organization, that is whether it is productive or destructive to the systems self-maintenance.[306] Whereas in Maturana cognition only is considered in its minimal sense, the interactions the living unit can enter into being constrained only by the basic requirement of self-continuance,[307] Thompson expands this fundamental relation and the organizational closure that demands it so that it

[304] Thompson (2007) page 13
[305] Ibid. page 124
[306] Walter (2009) page 106
[307] Maturana and Varela (1980) page 9-10

becomes a tool for understanding cognition all the way up to the rational human realm.[308]

4.2 Activity and the emergence of semantic networks

In the section on Juarrero I presented a downward causation construed as the constraintive powers of higher-order semantic organizing powers on bottom-up energetic and thermodynamic conditions, where the central element was the maintaining of identity by way of dynamic openness. The phase-space or ontogenic landscape of a dynamic system describes temporally extended patterns of activity where the flow of complex temporal structures mutually and simultaneously influence each other[309] providing more recursive coordination than the connectionist model of the mind. Rather than information-processing in terms of input-output relations we have a normatively guided top-down restrictive force that determine the semantic values of components that are to be integrated into this system.[310] The central point in my presentation of Juarrero's work is that closed systems can not be autonomous in the self-determining sense characterizing living systems, but must fundamentally be far-from-equilibrium systems that are constantly active in exchanging matter and energy with their surroundings. It is only by being in continuous interaction with some encompassing context that the higher-order properties come to be activated. Consequently, in the theory of autopoiesis any semantic or informational content of the living being does not come about by information gained through sensory input being processed and resulting in behavioural output, but is created through the reciprocal influences of the organizational closness (or identity) and its material openness.

Unlike previous theories of cognition and the metaphysics they inform, the systems that emerge from the accounts of Juarrero and Maturana are fundamentally interconnected with their surroundings and it is this mutual dependence between the

[308] "Cogniton is behaviour or conduct in relation to meaning and norms that the system itself enacts or brings forth on the basis of its autonomy" (Thompson (2007) page 126)
[309] Thompson (2007) page 43
[310] The autonomous organization holds between processes, not static entities.(Thompson (2007) page 44).

systems organizational closure and its performance in its structural coupling with its environment that allows it to be considered as an autonomous agent possessing mental capacities. Unlike the disembodied information-processing units described by earlier models of cognition the higher-order biological systems of dynamic theories are actively forming their environment in order to maintain the organizational structure that defines it as a system.[311] Instead of putting mind and body, or subjectivity and objectivity against each other Thompson believes that; "We see the co-emergence of inside and outside, of selfhood and a correlative world or environment of otherness, through the generic mechanism of network closure (autonomy) and its physical embodiment."[312] The distinction between the inner, subjective world of experience and the outer, objective world becomes void as they are not conceived of as pre-existing separate spheres, but "mutually specifying domains enacted or brought forth by the structural coupling of the system and its environment."[313] [314] Instead of the two being pitted against each other, the cognizing subject and the natural world become interdependent units and the Cartesian picture of the mind as 'inner' dissolves. There is no strict division between the inner syntactic properties of the brain and the externally related intentional or semantic properties of the mind because the two are essentially interdependent.

4.3 The reflexive downward causality of the cognizing cell

Thompson's project is to understand how the characteristics of mind and those of biological life can be understood in relation to one-another, and in line with his belief that there is a deep continuity between the two he begins at the bottom, by considering how this interdependence is present in the nature of the living cell.[315] Although the

[311] The domain of states available to the (nervous) system (as an operationally closed network) is a function of its history of interactions with the rest of the body and the environment.
[312] Thompson (2007) page 49
[313] Ibid. page 26
[314] Phenomenological perspectives paly a central role in Thompsons theory of embodied dynamicism; "central to the enactive approach is the idea that mind science and phenomenology can be linked in a reciprocal and mutually illuminating way- the enactive approach uses phenomenology to explicate mind science and visa versa" (Thompson (2007) page 265). Unfortunately it is outside the scope of this paper to consider these aspects.
[315] The cell was also the subject of interest for Maturana and Varela's exploration of autopoiesis; their theory was originally formulated to apply to productive forces in the molecular domain and they considered the living cell the paradigmatic case of autopoiesis (Thompson (2007) page 106)

cognitive powers of the cell are far removed from those of the rational human being, the fundamental mechanisms are the same. Insofar as cognition pertains to the behaviour of a system in relation to its environment the terms is as applicable here as in the human realm, although the meanings or norms it brings forth in this interaction have minimal content. The cell embodies a circular causality that is far from theoretical and that satisfies the three criteria of autopoietic organization. Thompson presents these criteria as a checklist that reflect the criteria Maturana put forth to determine whether a system can be counted as living or not.

The first is to determine whether the system in question has a semi-permeable boundary that allows for discrimination between the inside and the outside[316]**.** By way of continuous maintaining of its membrane the cell stands out as a unity against a chemical background, giving it its identity as an organizational unit. This boundary is not sealed, allowing for exchanges of matter and energy with the outside, thereby comprising its thermodynamic openness. So a cell meets the first criteria for being considered as a living system.[317] The second is whether the components of the boundary are being produced by a network of reactions that takes place within this boundary, not just on the boundary.[318] The whole of the internalized reaction-network must participate in this self-maintenance as it is in the metabolic network of the cell that the assimilation of select compounds from the environment takes place.[319] The cell constantly regenerates the components both of itself and the boundary by way of selectively importing outside elements, selected by their value as viability-sustaining components**.** If this internal process of production is interrupted, the cell will disperse and cease to be a self-contained unity.[320] The third point is to consider whether the two previous points are interdependent. If the internal reaction-network is dependent on the boundary it produces for its continued activity, then the system can rightly be called autopoietic and therefore living. The cells boundary, or more specifically its membrane, is determined by metabolic processes within the cell which are made possible by those same boundaries and so the two are interdependent. All this must of course take place in the physical space, as is the demand of living systems to be

[316] Thompson (2007) page 103
[317] Maturana and Varela (1980) page 94
[318] Thompson (2007) page 116
[319] Ibid. page 125
[320] Homeostatis is maintained as when the rate of disintegration of the membrane is balanced by the repair process (Thompson (2007) page 112).

defined autopoetically, and we see here the how cognitive interactions subserve the maintenance of organizational unity through metabolism.[321] In Maturana's words;

> What makes this system a unity with identity and individuality is that all the relations of production are coordinated in a system describable as an homeostatic system that has its own unitary character as the variable that it maintains constant through the production of its components.[322]

and further;

> This unity is, thus, a topological unity in the space in which the components have existence as entities that may interact and have relations. For living systems such a space is the physical space.[323]

Whereas Kim considered such reflexive downward causality to be incoherent, the self-maintaining activity of the living cell demonstrates exactly such a circular causation. It is by way of this continued self-production through the recognition of the fitness and importing of elements in its cognitive domain that the system persists as a spatially distinct individual,[324] and should this process of metabolic assimilation of select compounds from its environment stop it will thereby cease to exist as a distinct entity. In other worlds, it will die. Even here, the activity of relating to environmental perturbations in ways determined by meanings and norms brought forth on the basis of its autonomy[325], that is cognition, is indispensable in understanding its nature**.** Thompson says;

> Thus the cell embodies a circular process of self-generation; thanks to its metabolic network, it continually replaces components that are being destroyed, including the membrane, and this continually regenerates the difference between itself and everything else.[326]

---

[321] The circular definition of autopoiesis; that the system is organized in such a way that its constituent processes produce the very components necessary for the continuance of those same processes can be seen here*;* the metabolic processes within the cell determines its boundaries, but the metabolic processes themselves are made possible by those very boundaries (Thompson (2007) page 99).

[322] Maturana and Varela (1980) page 92
[323] Ibid. page 93-94
[324] Thompson (2007) page 101
[325] Ibid. page 126
[326] Thompson (2007) page 99

Here we have the minimal case of an entity which satisfies the requirements of autopoiesis and that therefore can be characterized as living. On the other hand, we have existents that fail to be counted as living because they do not satisfy the three criteria presented above. A borderline case is that of the virus. It does not produce its own boundary from within and it has no metabolism of its own by which to exchange matter and energy with the environment.[327] In short, it does not live up to the criteria of the living as autopoietic systems realized in the physical space.

The cell, being a minimal case of autopoiesis differs from those systems that are second-order in being composed of such first order autopoietic systems. The question is whether the criteria pertaining to cellular organization can be applied to such higher-order systems or if notions of boundary and internal reaction network must be interpreted so as to include only organization on this bottom level. In this case any system above this level would simply be aggregates of their first-order components, making them vulnerable to the arguments Kim presents against considering aggregates as in possession of distinct causal powers. Thompson believes that interpreting boundary to mean only semi-permeable membrane or skin is too narrow,[328] and in this he is in agreement with the biologists presented in a previous section who consider whole insects-societies autopoietic in the sense that they are self-organizing in relation to a domain of interaction. Keeping in line with the emphasizing of organization over physical structure, he says that; "Rather, the crucial matter is that the system produce and regulate its own internal topology and functional boundary, not the particular physical structure that realizes this boundary". [329] This means that second-order systems that consist of such first-order minimal autopoietic systems can be considered on the basis of the same criteria as these. What is missing in first-order autopoietic systems however is agency, which comes about through the emergence of sensory motor faculties that provides the organism with flexibility and adaptability in its metabolic maintenance. Without such capacities the organism cannot interact with its environment in an autonomic sense and intentional relations do not obtain.

[327] Thompson (2007) page 123
[328] Ibid. page 121
[329] Ibid. page 107

4.4 The intentionality of metabolism

Cognition in a cell-perspective is still a long way of the goal of understanding how the mental can exhibit causal powers in the realm of the rational human, but given the thesis that life and mind are fundamentally interconnected cognitive capacitates must be placed along an evolutionary continuum;

> What we humans experience as thinking, planning, deciding, and acting for reasons must have roots in similar but simpler forms of mental activity in lower (or less complex) organisms. If we can appreciate the expanding role of the mental in shaping behaviour as we move up levels of complexity within animal species, then the causal role of the mental in human life begins to appear less mysterious.[330]

Even as we climb the evolutionary ladder, cognition remains intimately tied to metabolism. It is the metabolism of the biological entity that fundamentally determines the internal norms that its activity or contextual interactions are regulated by[331] and it is by these that the elements contained in the domain of interactions become meaningful. The demands of metabolism determine the parameters of interaction because the internal metabolic reaction-network dictates the sustainability-requirements which its interactions must be subservient to.[332] It is through metabolism that the organism fundamentally maintains its identity; "Without incessant metabolic exchange with the world there can be no emancipation of dynamic selfhood from mere material persistence".[333]

Even in the minimal case of biological cells we can see that the demands of sustaining viability requires that the cell relates to its environment in a certain way, the importing of molecules from the environment determined by the viability-norms set by the metabolic demands of the system. In the biological cell the minimal requirements set by its metabolic demands gives it a primitive agency. Thompson provides the activity of bacterial (prokaryotic) cells as an example of how life as a process of cognition involves agency, albeight in a rudimentary sense. Moving about

[330] Murphy and Brown (2009) page 109

[331] Thompson (2007) page 153
[332] Ibid. page 126
[333] Thompson (2007) page 152

in the presence of a sucrose gradient they will tumble about until they have an orientation that will increase their exposure to this nutrient. They will then swim forward towards where the concentration of sucrose is at its greatest[334]. These basic organisms are capable of adapting their behavioural responses on the evaluation of sensory feedback in relation to the internal goal-states set by its metabolic demands[335] and modify the processes of exchange with the environment through adaptive responses. In the case of bacterium's ability to sense the concentration of sucrose in its immediate environment and move accordingly, the terms sensory response and action must be understood in a broad sense and does not indicate any awareness of behavioural flexibility. Here it is easier to see what is meant by the statement; "information is the intentional relation of the system to its milieu".[336] The meaning of the nutrients in the bacteria's environment does not contain this semantic quality apart from their being valued as relevant for the maintenance of the bacteria's identity or viability. The meanings of autonomous systems states are formed within the context of the systems dynamics and structural coupling"[337]

Francisco Varela tries to capture this dynamic interdependent pattern of activity apparent in the above example in his twin notions of identity and sense-making.[338] The first of the two concerns the production and maintenance of a dynamic or organizational identity in the face of material change, as captured in the term autopoiesis, while the second is intentionality in its minimal and biological form. It is the transformation of a physicochemical world into an environment of significance and value, in relation to the maintenance of the biological systems viability. While the first is logically prior to the second, identity can not be maintained without the intentional or sense making activity of cognition. The identity-making that autopoiesis entails establishes" logically and operationally the reference-point or perspective for sense-making and a domain of interactions"[339], so the content of the intentional relation of the organism to its environment is subservient to the maintenance of such

[334] Thompson (2007) page 74
[335] Murphy and Brown (2009) page 210
[336] Thompson (2007) page 59
[337] Ibid. page 58
[338] Here he distances himself from his earlier stance that norms (and purpose) do not pertain to the organization of the living as such, as was expressed in his and Maturanas original theory of autopoiesis.
[339] Thompson (2007) page 147

an identity, or the autopoietic organization. This notion of intentionality provides substance to Maturana's explication of cognition as the acting or behaving in the domain interactions defined by its autopoietic organization.[340] It supplies the organism with meaning and norms that are brought forth in the process of this interaction with its environment and as this domain of interaction increases, so does the richness of its semantical content. From the basic normative fitness-criteria of maintaining homeostasis through metabolic activity in the living cell to the immensely complex operations of belief and desire-guided human thought and behaviour, the common denominator is the relation of the system to its context, fundamented in thermodynamical openness and organizational closeness. In order to remain viable the autopoietic system has to make sense of this world, and on Varela's account of intentional sense-making this relation between the system and its domain of interactions is normative in a deeper sense than that suggested by his and Maturana's earlier account of the cognitive relations involved in self-maintenance. As seen in the example of the sense-making activity of prokaryotic cells, the emergence of the capacity of the organism to actively regulate itself with respects to its conditions of viability and thereby modify its environment according to the internal norms of its activity provides the graded norms that constrain this same activity[341]. The metabolic demands of the system fundamentally determines the values of the elements in its surroundings, such as the food-significance provided to sucrose as the bacteria metabolizes sucrose-molecules,[342] but as one goes up the scale of evolutionary complexity the all-or-nothing norm of basic self-continuance implied in the original theory of autopoiesis does no longer capture the cognitive capacities in operation. Although intentionality arises fundamentally from the operational closure and interactive dynamics of autopoiesis, the emergence of the nervous-system brings an added degree of agency for the organism and a corresponding increase in the richness of the meaningfulness and normativity of the relation between organism and environment. The nervous-system connects distant sensory-motor processes, subsuming them in operationally closed sensory-motor networks,[343] so increasing the

[340] Thompson (2007) page 124
[341] Ibid. page 148
[342] Ibid. page 154
[343] Ibid. page 159

capacity for adaptability in terms of behavioural flexibility.[344] Accordingly; "Living goes from being a process of basic cognition to being sense-making, of bringing forth significance and value".[345]

4.5 The emerging human mind

At the level of organisms with perceptual faculties the top-down constraintive capacities have, as the result of a greater degree of structural complexification of the brain,[346] intentional content far beyond that of the sucrose-seeking bacteria. This means more complex evaluative systems, greater amounts of information available about the environment and increasing behaviour flexibility, resulting in enhanced levels of self-direction or in Juarrero's terms; a greater degree of decoupling from lower-level forces. The capacity for environmental analysis and accompanying behaviour responses of the perceiving animal is dependent on the complexity of its nervous-system. In short, the level of adaptability and flexibility of the interrelationships between environmental feedback and behavioural responses is a function of evaluative capacities concerning the two. In insects the nervous-system is still rudimentary so any adaptability of behaviour is primarily brought about by genetic variations, but behaviour is nevertheless modulated on the basis of immediate feedback. There is a basic dynamic of action-feedback-evaluation-action loops present here but it is still minimal. The evaluation of behaviour has no learning or memory dimension and so the responses or behaviour flexibility is limited to evaluation of immediate environmental feedback.[347] With the introduction of memory the action-loops complexify further, as the animal is able to keep track of its previous behaviour and use this information to modulate ongoing behaviour. Here the downward power occurs as a successful pattern of behaviour (meaning that there is positive feedback from the environment in relation to a particular goal-state of the organism) causes particular brain-systems to become active in order to reinforce the specific neural pattern involved in the preceding behaviour.[348] To put it in terms of ontogenic

[344] "The capacity to change in relation to changing conditions in a viable (and not necessarily optimal) way" (Thompson (2007) page 159).

[345] Thompson (2007) page 157
[346] Ibid. page 104
[347] Murphy and Brown (2009) page 113
[348] Ibid. page 114-115

landscapes: as a particular course of action shows itself successful, the meaning or content that the network attributes to the external events satisfying the demands of its operational closeness takes the shape of or reinforces an attractor. As this strengthens through repeated reinforcing the system will tend toward this pathway, making a repeat of this behaviour more likely. Attractors are viewed here as activity-patterns of the brain,[349] the flow of which is constrained by the larger network they play a part in.

An even greater degree of behavioural flexibility accompanies the introduction of long term memory and the capacity for off-line reasoning. Here the cognitive domain and our sense-making capabilities does not concern just that which is accessible to us in the immediate physical world because, largely owning to the acquisition of language and the capacity for manipulating information off-line, the dynamics of sensory-motor activity become reflexive. This can be seen in higher-order animals as well as humans. An example is the problem-solving capacities of chimpanzees, who demonstrate the ability of representational manipulation in situations demanding such. When presented with a banana hanging out of reach and two poles, both too short to reach the banana, chimpanzees will first try to reach the food with either of the poles. Not being successful in this they will retreat and then, after a period of what appears to be the manipulation of mental images of the poles, the bananas and potential actions regarding the two, immediately proceed to put the poles together and pick down the bananas.[350] Here the semantic content of intentional acts are only secondary related to the immediate physiochemical environment and thus the domain of interactions of the system widens to include also that which is present only as a potential source of 'metabolic food'.

The capacity for off-line manipulations of representations and symbols means that we can set aside immediate goals for the attainment of more distant, and not immediately related to maintainance of metabolic efficacy. Related, normativity is no longer directly tied to the maintaining of homeostasis through metabolism but is complexified as we interact with social and cultural environments. A great deal of human behaviour is unconsciously intentional in that it proceeds automatically, meaning that we are not at the time aware of the off-line simulations taking place.

[349] Walter (2009) page 229
[350] Murphy and Brown (2009) page 124

Nevertheless, even though human action and accompanying thought is not always consciously intentional, our capacity for self-reflective awareness means that even acts that are not initially conscious can become the object of reflections or evaluations if it is needed. We have here a kind of third-order emergence of mental capacities which differ from the goal-directed intentionality of its precursors in that reflection can be turned on itself and articulated as reasons.

The capacity for self-reflective awareness and the flexibility of action provided by off-line 'mental imaging' appears to coincide with the expansion of the prefrontal cortex of the brain. That is not to say that human agency and adaptability is to be located here, because this area has extensive re-entrant, or two-way, interactions with all of the motor, sensory and affective areas of the cortex,[351] but it does seem to play an important role in the integration of perceptual information and modulation of behavioural response. The prefrontal cortex' role in organizing "goal-directed behaviour with respect to changing immediate and long-term context"[352] is exercised through top-down influences of lower-level control systems in the brain, constraining the freedom of these, all the while being dependent on the feed-forward of the activations of lower-levels which are in direct contact with environmental influences. Here is a model of the brain as an extremely rich version of a complex dynamic system, the prefrontal cortex being an important role-player in larger whole-brain networks that evaluate and modulate action by way of context-sensitive restraints[353] The evolvement of the large prefrontal cortex in humans allows for richer interactions and increased top-down regulation by higher order supervisory systems capable of evaluating lower-order cognitive processes.[354]

I stated previously that the emergence of capacitates for intentional sensory-motor interactions with a co-emerging environment should be understood as lying along an evolutionary continuum. There is however one major 'jump' in this continuity that takes place as the capacity for symbolic language emerges, in concert with the development of the prefrontal cortex. Here it is the mapping of relationships among words, not the relationship between the symbol and the external environment (as it is

[351] Ibid. page 132
[352] Murphy and Brown (2009) page 135
[353] Ibid. page 160
[354] Ibid. page 223-224

in simpler forms of mental images) that provides cognitive efficacy.[355] Understanding the causal power of language as that of contextual constraints on the neural activities that embody its use means recognizing the organization of the brain as a nested hierarchy of top-down controls through "recurrent connections that enable re-entrant feedback loops".[356] In brain-terms, feedback from the environment serves to as a selective influence on the vast number of neural connections so as to strengthen those that are conducive to its performance. The brain becomes a self-modifying system, where the top-down constraints of the semantic networks modify its neural structure in response to the norms of its operations. "The organizing system is trained by the environment, and it in turn has downward casual efficacy in governing lower-level cognitive processes and thus the neural structures that subserve them".[357] The formation of abstract concepts takes place as the development of broader basins of attraction in semantic space, originating out of our sensory-motor experiences. Complex meanings and concepts arise out of the basic requirement of autopoietic cognition, language being only intelligible within a context of action.[358]

The issue of language is complex and I cannot go further into it here, so I will restrict myself to saying that it is with the emergence of language that we reach the pinnacle of rational capacities, the ability to form and act according to explicit reasons. We seem here to have reached the kind of intentionality that philosophers of mind are concerned with. Unfortunately there is no room here for exploring further the specific mental capacities making up the human mind, of which consciousness (as the self-reflective awareness of such capacities) is of special interest, but hopefully what has been presented here provides a general outline of how such capacities can be understood.

[355] Ibid. page 126-127
[356] Ibid. page 172
[357] Murphy and Brown (2009) page 231

[358] "Most of our conceptual resources come from metaphorical extention of concepts derived directly from sensory-motor experiences" (Murphy and Brown (2009) page 213)

4.6 Dynamic action from cells to humans

I have attempted to trace intentionality, or more generally the activity of cognition, from its minimal form in cells and bacteria to the deliberative off-line reasoning of language-using humans**,** the common denominator being that the organism is normatively related to an outside specified by the self-production of an inside.[359] In autopoietic terms; "the dynamics of the system is characterized by an invariant topological pattern that is recursively produced by the system and that defines an outside to which the system is actively and normatively related."[360]

What sets this understanding of cognition apart from the earlier presented connectionist models is that the neural processes they are constituted by are embedded in larger causal and semantic networks, these being contextualized physical states. Meaning is not in the brain and not independent of the embeddeness of the cognizing being in environmental, social and cultural contexts.[361] Being a bearer of information (in the sense of having semantic content) has causal power through its role in action-feedback loops, not through identification with a certain area of neuronal activity, although the ability for adaptive responses is instantiated as structural changes in these networks. To put it in the words of dynamic system theory; intentional action is the continuous top-down control of contextual constraints on the probability distribution of behavioural alternatives. In terms of ontogenic landscapes; "feedback between external circumstances and internal dynamics can drive neural dynamics far enough from equilibrium so that one attractor becomes an intention, reorganizing the landscape."[362] The role of the mental is as a structuring influence on the flow of events on lower levels.

[359] Thompson (2007) page 163
[360] Ibid. page 260

[361] "the meaning and content of beliefs and intentions derive from their relations" (Murphy and Brown (2009) page 222)

[362] Murphy and Brown (2007) page 228

The interdependence of mind and life means that our mental lives cannot be understood apart from our bodily activity, the most fundamental of these being the basic self-regulation of the living cell. In organisms with a nervous-system the domain of activity and thereby the richness of content and intentionality expand as the autonomous sensory-motor agent emerges. This kind of sensory-motor coupling with the domain of activity in turn lays the fundament for the evolvement of a mental life in terms of beliefs and desires, and ultimately reflective self-awareness. It is permissible I believe, to claim that the appearance of substantial mental life coincides with the appearance of the faculty of perception. It may not be the self-aware consciousness of the rational human being, but it is here that the issues of agency and cognitive capacities start to become relevant. On the account presented by the enactive approach, cognition, as the intentional activity of producing and maintaining the self by interaction with an environment, has powers that can not be sub-seeded by the intrinsic material composition of the organisms as is Kim's contention, because the parts making up the whole are only definable in relation to the whole it is a part of.

I stated in the beginning of this paper that the intentional nature of the mental was, along with its related normativity, a great part of what apparently makes it difficult to fit into a physical world conceived of as consisting of particles with determinate spatio-temporal locality. The consequence of the dynamic approach to mind-body relations is that mental life is fundamentally interwoven with our identity-maintaining activity of relating to a physiochemical world, and so cognition cannot be understood apart from action. Mental influences do not intervening in a linear causal fashion between sensory 'input' and motor 'output' as Kim's functional approach suggests, they are the structuring powers of the organism in action:

> In other words, as a skilful activity of the whole animal or person, perceptual experience emerges from the continuous and reciprocal (non-linear) interactions of sensory, motor, and cognitive processes, and is thereby constituted by motor behaviour, sensory stimulation, and practical knowledge. On the brain side, neural states are described not at the level of their intrinsic neurophysiological properties or as mere neural correlates of mental states, but rather in terms of how they participate in dynamic sensory-motor patterns involving the whole active organism.[363]

[363] Thompson (2007) page 256-257

The agency and related selfhood which theories of mental causation seek to capture is already present in the activity of the sensory-motor organism, originating in the first-order autopoietic organization of the living cell. There this circular organization and its accompanying coupling with the environment was regulated by a minimal viability-constraint, whereas the richness of semantic content in the sphere of human rationality comes about the emergence of second and third-order organizational powers which provide us with domains of interactions far beyond that which is set by the demands of metabolism. Cognition understood in this sense is accordingly essentially tied to our activity as living beings, right down to its beginnings as minimal life and can therefore only be understood in terms of the criteria that are common to all those existents that can claim sentient existence. The whole is thereby not the sum of its intrinsic parts and their relations, as is Kim's assumption, but rather the organizational closure of its parts that gives all living systems their unity.[364]

SECTION FIVE: Summary and conclusions

If physicalism is to survive as a worldview for us, it must show just where we belong in a physical world, and this means that it must give an account of our status as conscious creatures with powers to affect our surroundings in virtue of our consciousness and mentality.[365]

I began this paper by expressing the issue to be treated as a question of how there can be such a thing as mind in a world consisting ultimately of nothing but bits of matter distributed over space-time, behaving in accordance with physical law and by this presented the dominant physicalist parameter that the mind-body discussion is framed by. Jaegwon Kim's belief that all higher-order existents must be reducible to the entities treated by fundamental physics in order to be allowed within the physical domain was shown to be a consequence of this physicalist paradigm. Otherwise, he

[364] Thompson (2007) page 449
[365] Kim (2005) page 30

and his cohort contend, they can not claim any causal standing in a world that consists of nothing but matter. Very briefly stated, the argument went like this; each physical event, such as that of initiating an action has a complete and sufficient physical cause, and if it has a physical cause it can not have a mental cause in addition (so the causal exclusion-principle says). Thus mental events must be reduced to physical events or they have no causal influence in the world and are therefore not real (on account of the equation of existence and causal power as stated in Alexander's dictum). If, as assumed, there is a relation of supervenience between mental and physical events, then all the casual work necessarily goes on at the bottom level and any impression of vertically directed causality is merely on account of the dependence of the supervenient event or property instantiation on its subvenient base. As a consequence of this Kim believes that there are only two options available to the physicalist; finding a satisfying way of reducing the mental to the physical or denying the existence of the mental altogether. Otherwise we are going to have to go for dualism and thereby postulate an altogether non-physical domain that somehow mysteriously comes to intervene in the goings-on in the physical sphere**.**

The discussion is often portrayed as being between reductionists of Kim's sort and property-dualists, who posit distinct properties standing in opposition to each other, but what I have tried to suggest by presenting theories that consider the defining feature of the living to be exercise of genuinely emergent properties of self-organization, suggesting an either-or version of the causal powers of the mental does not do capture the graded complexity of living organisms as physical entities. The proponents of a conception of mentality that does not reduce to the physical in the way it is construed by Kim, for whom the notion of the physical rests on the explanatory completeness of physics, do not in any way wish to postulate entities that are non-physical. They still consider themselves physicalists but distance themselves from the kind of physicalism that states that all that is worth scientifically contemplating are bits of matter and their aggregates behaving in accordance with the laws of microphysics. That is not to say that the forces operating at the levels described by physics are irrelevant but that they alone cannot sufficiently account for all the phenomena that deserve scientifically respectable explanations, there being physical forces that are not adequately described by the laws governing their basic

parts alone. Rather, macro-properties exhibited by living systems play a causal role for which the highly specific micro-based macro-properties of Kim is unsuited.[366] Minds cannot simply be reduced to brains because the mind is essentially related to the greater system that the brain is embedded in. The non-reductionist treatment of the matter as presented in this paper does not rest on considering mainly the metaphysical or analytical aspects of the concepts involved but rather on what features are common to all that which can be characterized as living, viewed in an empirical light. The basic tenet underlying these theories is that there is a continuity of life and mind, the idea that organisms even at its lowest forms portend mind and that mind remains fundamentally part of the organic even at its highest reaches.[367]

To fully grasp complex systems such as biological existents, and ultimately the human organism, we must consider forces characterized by other elements than the mere energetic, but that can still be incorporated into a physical universe in a satisfactory way. As I have attempted to show in presenting Alicia Juarrero's account of the downward organizational forces that operate in open systems far from equilibrium, such higher-order influences on microphysical particulars is a common denominator of all systems that are rightly termed complex. There is no mystery concerning how these powers operate, their being fully explainable in scientifically respectable terms and they do not compete with lower-level laws, but rather supplement them. This challenges Kim's claim that such circular or recurrent causality is not possible to conceive of in a coherent manner.

At the basis lies a disagreement over the understanding of the dependence-relation obtaining between higher-order properties and the micro-physical particulars that realize them. Kim expresses this dependence through the notion of mereological supervenience, which has three central characteristics. Firstly, the existents that are related to each other here are static elements, instantiations of properties at a certain point in time that are always synchronous with those that are co-instantiated. Inspired by the dominant views of cognitive science the mental is interpreted as supervening on the internal processing of information by neuronal units. The determining relation

[366] Strand (2007) page 64
[367] Thompson (2007) page 128

between these is its synchronous microstructure,[368] so that capturing the total inner state of the brain at a given point in time is sufficient to give a full description of the mental. This leads to the second point, that all such supervening events are reductively describable in terms of their microstructures. That is, they are decomposable into the properties of and relations among their basic particles. Nothing more is needed to describe the properties of the whole other than the intrinsic properties of their parts, because as mereological supervenience states; "The fact that s has P is fixed once the micro-constituents of s and the properties and relations characterizing these constituents are fixed."[369] This means that the relation between higher-order properties and their realizers necessarily is asymmetric, which is the third characteristic of such supervenience. The dependence-relation is one way; the parts making up the base-level determine the whole of the supervenient level without any such determination of the parts by the whole they make up. These three elements are precisely what make it so hard to fit the mental within a physicalist scheme. While the properties of the mental states that are allowed causal powers as conceived of within the physicalist paradigm underlying mereological supervenience are expected to be local and intrinsic, the contents of mental states are essentially relational.[370] This leads to Kim's question; "How can extrinsic mental properties be causally efficacious in behaviour causation?"[371] As long as all causality is conceived of as only taking place among the micro-structural components of events supervenient on a causally closed micro level, then there is no way to fit mentality in without stripping it of its essential characteristics.

The non-reductive theories presented in this paper believe that the whole should not be equated with the sum of its parts, but rather with the organizational closure of its parts. This means going against the assumptions underlying the concept of supervenience as it is understood by Kim. Whereas classical physics, whose principles underlie Kim's approach to reduction were mainly concerned mainly with closed systems that are unproblematically describable through the three conditions mentioned above, living systems are fundamentally open systems. This means that the

[368] Kim (2005) page 36
[369] Kim (1998) page 85
[370] Kim (2008) page 453
[371] Ibid.

content of their material composition is secondary to invariant topological patterns in defining their identity. In combination with this organizational unity, openness means that such systems can maintain a stable identity despite of constant change in its compositional matter. It is precisely the capacity of the system to actively relate to its contextual conditions that facilitates the emergence of life and of cognition.[372] The internal composition and relations making up the organism does not alone fully capture its nature as living, because in addition the organizational unity plays an indispensable part in the structuring of the activity of these material components. Accordingly, this organizational unity and the material composition instantiating it are interdependent, embodying a circularity that the relation of supervenience does not allow. Donald Campbell puts it this way;

> supervenience is explicated in terms of entities-particles-properties. This is basically an Aristotelian metaphysics, and is an inadequate metaphysics for relationships and process, most specially open processes. "Entities" that are organizations of underlying far-from-equilibrium process are not supervenient as long as supervenience discounts external relations, and so long as it only counts lower level constituents as part of the supervenience-base.[373]

In the dynamic approach presented in this paper, cognition is viewed as extending beyond the agents' brain, the cognizing beings nervous system, body and its environment being viewed as coupled dynamical systems. The role of cognition as the activity of maintaining organizational unity was expressed through Maturana and Varela's theory of autopoiesis, where the notion of cognition was introduced as essentially a relation between an organism as a self-organizing system and the context of interactions it operates within. The combination of Juarrero's openness and Maturana's cognition suggests an alternative picture of the relation between higher-order processes and their realizers. Descriptions of the internal state of a being at a given time and its representational content is replaced with a focus is on the unfolding trajectory of the system's state and the internal and external forces that shape this trajectory. Cognition is presented as fundamentally a process of relation and not primarily as the state of an organism. This counters the second point of Kim's supervenience-picture, namely that all higher-order properties are decomposable into

[372] As mentioned earlier, without this capacity for change the organism will die.
[373] Andersen et.al (2000) page 334

the properties of its parts. On the picture presented by the dynamic approach two apparently identical intrinsic states can have different semantic content, depending on how they are entrained into a network of identity-maintaining activity. The causal relations at play here are not limited to the horizontal causal chains occurring at the level of micro-physical components, because the relative strengths of these component processes are continuously affected by how they are retained into a organizational whole.

5.1 On the intentional nature of the constraintive forces

Understanding the nature of this circularity of self-production is essential to capturing the role of the mental. We have here emergent downward causation, a specific kind of reflexive global-to-local influence that happens in a system that has dynamic global coherence in and through collective self-organization.[374] It is through this circularity that the characteristics of mentality emerge and are explained**.** On Kim's account, causal processes are only physically respectable if they are reductively identifiable with events taking place at the level of their micro-particulars, as seen in the identification of functional mental properties with their first-order realizers as micro-structural properties based at the fundamental physical level. There cannot be any physically respectable forces here that are not reducible to the efficient causality of a postulated micro-physical level. The accusation against Kim was that this leaves the mental with no causal power of its own, to which his answer was that unless the non-reductionist could point out the mechanisms grounding such forces or powers, his approach was the best one there could be. The dynamic systems approach to the causal powers of the mental provides such an account, taking configurational forces or powers that are empirically demonstrable in physical systems and using them as tools to explain how causal forces of an intentional or normative nature are required to fully explain the operations of the living being. The active nature of the biological entity is emphasised more here than in connectionist approaches to cognition, where cognition is viewed as the processing of context-independent information encoded by

[374] Thompson (2007) page 434

internal structures. In the dynamic approach information is essentially created in the intentional relation of the system to its environment, determined by the systems operationally closed dynamics and its modes of structural coupling with the environment.

Summarizing this approach, the belief here is that the process of cognition by which a system maintains its self-organization by relation to its contextual conditions is structured by recurrent feedback-loops where the semantic information or (intentional 'aboutness') emerges from this essential activity of structural coupling. In living organisms with perceptual capacities this constant interaction takes the form of action-feedback-evaluation-action loops, where sensory responses guide action and action has consequences for subsequent sensory stimulation, subject to the constraint that systems maintain its viability. This does not take place as links in causal chains leading to physical behaviour, such as is the view of Kim[375] and of cognitive science portraying the mind as starting with sensory input and ending in behavioural output but as self-organizing forces modifying the flow of neuronal events.

The basic autopoietic property of generating own boundary and internal topology is there all the way, but is immensely complexified on the way to human beings. The downward constraintive power of the higher-order intentional properties of the active organism increases as the richness of intentional (or semantic) content increases. This process of increasing complexification cumulates in the specifically human capacity of acting for reasons and in the power of self-reflective consciousness.

It is important to recognize that the local efficient causes described by physics and the global-to-local structuring influences that constrain them are not to competing causes here in the way Kim's supervenience argument assumes. Tracing the mental and herein intentionality back to the fundamental defining characteristics of the living unit means that reasons do not need to be viewed as in opposition to causes. They are causes of a different kind, being constraintive rather than productive and not intrinsic to the system. But, as already mentioned, they play an indispensable part in the activity of the living organism which can never be captured by considering the physical parts of the system alone because; "actions resulting from reasons exhibit a

[375] "Functional properties are specified in terms of their roles as causal intermediaries between sensory inputs and behavioural outputs" (Kim (1998) page 19)

semantic dimension that physical processes lack."[376] It is not only actions resulting from explicit reasons that have this semantic dimension, but rather all interaction between a living system and its environment. The process of living necessarily involves normative selective criteria which in turn involves a cognitive relation between the organism and its context of operations.

5.2 What about closure and exclusion?

If it is the case that the intentional or normative aspects of human nature can be satisfactory explained naturalistically and without violating physicalism, what are we then to make of the principles that Kim considers to be defining of physicalism? Must the ideas of a causally closed universe and of causal exclusion be discarded if such semantic causes, as we might term them, are to claim in pass into the physical universe? As I elaborated in the section where the principle was presented, the demand of causal closure of the physical domain is intimately tied to a certain conception of the nature of the physical universe. The 'bottom level' making up the supervenience base to which all physical phenomena must be reducible is in Kim's view the elementary particles included in the closed physical domain of microphysics.

He does not however offer any explicit treatment of this principle of closure and the charge is, in Thompson's words, that;

> This position claims science for its support, but it is metaphysical in the sense of going beyond anything science itself tells us. I see no good reason to believe in such a thing as an "emergence base", where this means a configuration of pre-existing microphysical entities with intrinsic properties and causal powers that ground the macro-physical level. This image of nature as a mereologically ordered hierarchy grounded on a base level of particulars is a metaphysical principle projected onto science, whereas the image science project is of networks of processes at various spatiotemporal scales, with no base-level particulars that "upwardly" determine everything else.[377]

[376] Walter (2009) page 194
[377] Thompson (2007) page 439

In fact, the conservation-principle that underlies his understanding of causal closure need not be violated by taking configural or 'higher-level' powers into account. As stated when this principle was introduced, every physical system is either conservative or part of a larger system that is conservative. The domain of interactions that an open, far-from-equilibrium system is embedded within is to be understood as an encompassing system because, as Juarrero states; "when components are dynamically coupled and coordinated, a self-organizing network of components and its environment are in fact one system."[378] Accordingly, these interactions can affect the distribution of the amount of energy or momentum within the system in question. This seems to be in line with Juarrero's view of context-sensitive constraints as configurational forces affecting the probability-distribution of a system's constituents. Considering the energetic and thermodynamic requirments for basic autonomy means describing this downward force as the capacity of the system to manage the flow of matter and energy in order to maintain itself as a unity.[379]

All changes here are basically explicable in terms of energy or momentum-redistribution, but in addition to the instances of properties at the fundamental level there are forces constraining these constituent interactions in order to maintain a stable pattern amongst the constant turnover of material components. In terms of neural events, this causal force can be conceived of as modifying the flow of neuronal activation so as to make "the fact that a causal tree of neural events converge upon a particular bodily movement non-coincidental."[380] [381] The continuous affect of context and historicity on the relative strengths of component processes is hopefully sufficiently described throughout this paper, as well as the point that the micro-physical particulars not necessarily can claim ontological primacy in this scheme.

Apart from the likelyhood that the conservation of energy principle allows an encompassing system to affect the distribution of events of the system it contains, the

[378] Juarrero (2009) page 86
[379] Thompson (2007) page 46
[380] Gibb (2010) page371
[381] Considering that there are more potential synaptical connections in a single brain than there there are atoms in the entire known universe, the idea of forces 'steering' their activation towards some causal effect seems to play an indispensable explanatory role.

fact that Kim describes the causally closed domain as a bottom level of fundamental particles[382] is hardly in line with the findings of contemporary physics. Presenting the purposed explanatory completeness of physics as the doctrine underlying his argument for the reducibility of all higher-order causality to its domain should go hand in hand with a presentation of which principles of physics specifically he adheres to[383]. As long as these principles are kept wholly implicit the argument for the reducibility of higher-order existents remains incomplete.

Regarding the exclusion-principle, there seem to be no special problems of mental causation conceived of as downward constraintive forces of a complex system in relation to its contextual interactions, because of the fundamental interdependence between the different levels of causal existents. The underlying firing of neural assemblies does not explain sufficiently why a certain action was initiated in an organism with a nervous-system, because in addition the restraining forces of semantical context they are embedded within plays an indispensable part in making things happen as they do. One could of course question whether such downward power deserves the label causality at all, but if we consider causality in the common-sense way as that of having the power to affect the course of events, then it certainly does. In fact, it gives us an extra dimension, and are we to accept the idea that living organisms are not first and foremost concrete things but structures of processes, the notion of semantic forces constraining the flow of lower-level events is better suited to explain how these processes unfold than the view of mental operations as internal manipulations or processing of independent information.

Whereas the gist of the supervenience-argument was that considering distinct mental events in addition to physical events as sufficient causes of subsequent physical events left the mental cause redundant, the non-reductionist theories presented in this paper present the two causes as essentially interdependent. Summarizing, there are two different kinds of causes at play here, Kim's energetic or productive causes defined in terms of fundamental physical science, and the "negative" restraintive

[382] Kim (2005) page 66

[383] Given that the requirment for properties to be termed physical is that they are dealt with in fundamental physics or reducible to fundamental physical properties (Kim (2005) page 157)

causes of complex adaptive systems characterized by feedback processes from the activity of the organism, in which the product of the process is necessary for the process itself. While the first is intrinsic to the particulars of the system in question, the second is extrinsic in the sense that it is defined as the intentional relation of the system to its context of operations (environment), and neither can operate without the other. Both are however physically respectable in the sense that they do not resort to empirically unverifiable 'vital entelechies' and both are indispensable in understanding the operations of mind.

5.3 Epilogue

Kim sets the following requirements for the discarding of a framework;

> To motivate the discarding of a framework, we need independent reasons- we should be able to show it to be deficient, incomplete, or flawed in some fundamental way, independently of the fact that it generates puzzles and problems that we are unable to deal with.[384]

In my view, the dynamic approach offers better solutions to the problems of mind-body relations, which apparently arise from the tension between "the objective world of physical existence and the subjective world of experience",[385] as Kim puts it. It does so by addressing a deficiency in the model of physicalism espoused by Kim, namely the belief that the constitution of complex systems and herein the causal relations structuring this, can be adequately defined intrinsically, either microstructurally or functionally, while the mental characteristics must, in virtue of their intentional nature, be defined relationally in an extrinsic sense. Simply stated; cognizing beings such as us cannot be understood apart from the contexts of our interactions with the world, and that's where the supervenience-argument goes wrong. "You are a living body subject of experience and an intersubjective mental being",[386] Thompson's states, and it is this failure to recognize that the mind cannot be isolated from its environment that explains the difficulties of reductive physicalists approaches

[384] Kim (2005) page 30
[385] Ibid.
[386] Thompson (2007) page 243

to mental in describing how the mind operates in a physical world in a satisfactory way. The issue should not stated as a question of mind-brain relations but must be expanded to include the context of operations of the mind and brain, these being both body, physiochemical environment and, concerning intentional beings at the highest level, social and cultural contexts.

In closing I would like to point out that the approach to mentality and its causal powers presented here in response to Kim's challenge might suggest answers to some of the most persistent issues in philosophical discussion. Recognizing how our environmental, social and cultural contexts affects the way we think might spur us to supplement the conceptual analysis that philosophers are so fond of with a critical examination of how these concepts come to be. This requires cooperation between empirical science and philosophy and a willingness on the part of philosophers to take scientific findings into consideration. The philosophical debate on the mind-body issue as in any other should be sufficiently scientifically informed as; "the mental does not exist in a theoretical vacuum."[387] Science should as well be respondent to philosophically grounded critique of methodology and subject-matter, as awareness of the paradigms research is conducted within could provide those concerned with a particular problem in a certain domain of science with a wider perspective that can be conducive to progress. For example, a shift in methodology from focusing on isolating particular systems and studying what takes place within them to recognizing the problem of initial conditions and on context-relationalness could provide science with explanatory tools for solving problems that persist within the reductionist approach. As seen in the domain of biology, such a paradigm-shift is already underway. While Kim exemplifies the suspicion that downward causation is viewed with in philosophy, science commonly employs it as an explanatory tool.

In cognitive science specifically, abandoning the search for the seat of consciousness in favour of the study of the mind/brain as an extended and socially, environmentally and culturally embedded system, might prove to be a fruitful way to approach the central issues of this scientific domain. Although the reductionist research-strategy

[387] Walter (2009) page x

has proven useful in many areas; ”reductionism fails when confronted with the mind-brain problem- thinking changes the operational patterns of neural activities in my brain.”[388] The notion that intentionality exercises normative restraintive power on the lower order causal pathways gives the mental genuine causal efficacy, and it is in this downward efficacy on the arrangements of the compositional details of its realizers in the brain and that the following should be understood; “the whole is more than the sum of its parts, not in an ultimate, metaphysical sense, but in the important pragmatic sense that, given the properties of the parts and the laws of their interactions, it is not a trivial matter to infer the properties of the whole.”[389] The laws and relations governing what is commonly considered the physical domain must be supplemented by a description of the contextual relations structuring such physical events into complex systems such as ourselves.

In the end, what we have is that which the early emergentists sought- an approach that offers a middle way between the extremes of all-out reductionism and inconceivable dualism; not saving mind as separate from body, or making body all mind. This middle-way might prove the best way forward on the central issues in the discussion of how the mind and the physical world fit together.

[388] Ayala (1974) page xii
[389] Skolimowski (1974) page 211

BIBLIOGRAPHY

Andersen et.al (eds.) (2000): *Downward Causation. Minds, Bodies and Matter.* Aarhus: Aarhus University Press.

Atkinson, Rita L. et.al (eds.) (2000): *Hilgard's Introduction to Psychology,* Harcourt College Publishers.

Ayala, Francisco Jose & Dobzhansky, Theodosius (eds.) (1974): *Studies in the Philosophy of Biology. Reduction and Related Problems,* London: MacMillian Press.

Beckermann, A., Flohr, H. & Kim, J. (eds.) (1992): *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism,* Berlin: Walter de Gruyter.

Beckermann, Ansgar (1992): "Supervenience, Emergence, and Reduction", in Beckermann, Flohr & Kim (1992).

Beer, Randall D. (2000): "Dynamical approaches to cognitive science", *Trends in Cognitive Sciences –Vol.4 No.3 March 2000.* 91-99.

Beer, Stafford (1980): Preface to 'Autopoiesis", in Maturana and Varela (1980)

Block, Ned (2003): "Do Causal Powers Drain Away?", *Philosophy and Phenomenological Research, Vol.67(1) 2003.*133-150

Campbell, Donald T.(1974): "'Downward Causation in Hierarchically Organised Biological Systems", in Ayala & Dobzhansky (1974).

Campbell, Donald T.(1974):"Unjustified Variation and Selective Retention in Scientific Discovery", in Ayala & Dobzhansky (1974).

Campbell, Donald T. (2000): "Emergence", in Andersen et.al. (2000).

Gibb, Sophie (2010): "Closure Principles and the Laws of Conservation of Energy and Momentum", *Dialectica, Vol.64, No.3,* 363-384.

Globus et.al (eds.) (1976): *Consciousness and the Brain. A Scientific and Philosophical Inquiry,* New York: Plenum Press.

Juarrero, Alicia (1999). *Dynamics in Action: Intentional Behaviour as a Complex System,* Cambridge Massachusetts: MIT Press.

Juarrero, Alicia (2009): "Top-Down Causation and Autonomy in Complex Systems", in Murphy et.al (2009b).

Kim, Jaegwon (1989): "The Myth of Nonreductive Materialism", *Proceedings and Addresses of the American Philosophical Association* 63, 31-47.

Kim, Jaegwon (1992): ""Downward causation" in Emergentism and Non-reductive Physicalism", in Beckermann et.al (1992).

Kim, Jaegwon (1998): *Mind in a Physical World: An Essay on the Mind-Body problem,* Cambridge Massachusetts: MIT Press.

Kim, Jaegwon (1999): "Making Sense of Emergence", *Philosophical Studies* 95, 3-36.

Kim, Jaegwon (2005): *Physicalism, or Something Near Enough,* Princeton New Jersey: Princeton University Press.

Kim, Jaegwon (2008): "Mental Causation", in Lycan & Prinz (2008).

Lycan, William G. & Prinz, Jesse J. (eds.) (2008): *Mind and Cognition. An Anthology,* Malden Massachusetts: Blackwell Publishing.

Maturana, Humberto R. & Varela, Francisco J. (1980): *Autopoiesis and Cognition. The Realization of the Living,* London England: D. Reidel Publishing Company.

McLaughlin, Brian P. (1992): "The Rise and Fall of British Emergentism", in Beckermann, Flohr & Kim (1992).

Murphy, Nancey & Brown, Warren S. (2009): *Did My Neurons Make Me Do It?,* Oxford New York: Oxford University Press.

Murphy, Nancey et.al (eds) (2009b): *Downward Causation and the Neurobiology of Free Will,* Berlin: Springer-Verlag.

Skolimowski, Henryk (1974): "Problems of Rationality in Biology", in Ayala & Dobzhansky (1974).

Stephan, Achim (1992): "Emergence- A Systematic View of its Hisotrical Facets" in Beckermann, Flohr & Kim (1992).

Strand, Anders (2007): *The Metaphysics of Mental Causation*, University of Oslo: Ph.D Thesis in Philosophy.

Thompson, Evan (2007): *Mind In Life. Biology, Phenomenology, and the Sciences of Mind*, Cambridge Massachusetts: Harvard University Press.

Walter, Henrik (2009): *Neurophilosophy of Free Will,* Cambridge Massachusetts: MIT Press.

Weimer, Walter B.(1976): "Manifestations of Mind: Some Conceptual and Empirical Issues, in Globus et.al (1976).